The Annals of Statistics
2022, Vol. 50, No. 4, 2205–2230
https://doi.org/10.1214/22-AOS2183

ASYMPTOTIC INDEPENDENCE OF SPIKED EIGENVALUES AND LINEAR SPECTRAL STATISTICS FOR LARGE SAMPLE COVARIANCE MATRICES

BY ZHIXIANG ZHANG[1,a], SHURONG ZHENG[2,c], GUANGMING PAN[1,b] AND PING-SHOU ZHONG[3,d]

[1]*School of Physical and Mathematical Sciences, Nanyang Technological University,* [a]*S160011@e.ntu.edu.sg,* [b]*GMPAN@ntu.edu.sg*

[2]*School of Mathematics and Statistics, Northeast Normal University,* [c]*zhengsr@nenu.edu.cn*

[3]*Department of Mathematics, Statistics, and Computer Science, University of Illinois Chicago,* [d]*pszhong@uic.edu*

We consider general high-dimensional spiked sample covariance models and show that their leading sample spiked eigenvalues and their linear spectral statistics are asymptotically independent when the sample size and dimension are proportional to each other. As a byproduct, we also establish the central limit theorem of the leading sample spiked eigenvalues by removing the block diagonal assumption on the population covariance matrix, which is commonly needed in the literature. Moreover, we propose consistent estimators of the L_4 norm of the spiked population eigenvectors. Based on these results, we develop a new statistic to test the equality of two spiked population covariance matrices. Numerical studies show that the new test procedure is more powerful than some existing methods.

1. Introduction. Sample covariance matrices play a fundamental role in traditional multivariate statistics (see [1]). There has also been a significant interest in studying the asymptotic properties of the eigenvalues and eigenvectors of the sample covariance matrix in the high-dimensional setting where the data dimension p grows with the sample size n. These asymptotic properties have been used to make statistical inference, such as hypothesis testing or parameter estimation, about the population covariance matrices of high-dimensional data. Random matrix theory turns out to be a powerful tool for studying such asymptotic properties. One can refer to the monograph of [3] or the review paper of [27] for a comprehensive review.

The spiked covariance model appears naturally in the areas of wireless communication, speech recognition, genomics and genetics, finance, etc. It refers to a phenomenon that if the largest population eigenvalue is greater than some critical value then the largest sample eigenvalue will jump outside the bulk spectrum of the corresponding sample covariance matrix. Such a phenomenon has received much attention recently. The pioneer work of [18] considered a special spiked model with a $p \times p$ diagonal population covariance matrix

$$\boldsymbol{\Sigma} = \operatorname{diag}(\alpha_1, \dots, \alpha_K, 1, \dots, 1), \tag{1.1}$$

where $\alpha_1 \geq \cdots \geq \alpha_K > 1$ are referred to as spikes and $K < \infty$ is the number of spikes. [11] investigated the almost sure limits of the largest eigenvalues which depend on the critical value $1 + \sqrt{\gamma}$ when $p/n \to \gamma > 0$. [26] established the central limit theorem for the spiked sample eigenvalues under the Gaussian assumption on the data. [4] extended Paul's results by removing the Gaussian assumption, but assumed a block diagonal structure on the population covariance matrix. [5] further generalized the spiked model by considering an arbitrary nonspiked part of the population covariance matrix instead of identity (but still a block diag-

Received August 2020; revised February 2022.
MSC2020 subject classifications. Primary 62E20; secondary 62H15.
Key words and phrases. Leading spiked eigenvalues, sample covariance matrix, linear spectral statistics, central limit theorem.

onal structure). They classified the spikes into distant spikes and close spikes, and discussed the almost sure limits for two types of spikes and established a central limit theorem for distant spikes. [16] obtained the CLT for the spiked eigenvalues by relaxing the block diagonal structure on the population covariance matrix. They assumed that either the largest entries of the population eigenvectors corresponding to the spikes tend to zero or the fourth moment of the underlying variables must match with that of Gaussian distributions. After our manuscript was posted on arXiv and in submission, we were informed by one of the authors that they submitted a new manuscript [17] by further removing the fourth moment matching condition. In addition to the above literature about the bounded spikes, we would also like to mention that there is some literature about the unbounded spikes and one may see [14] and the reference therein.

The study of linear spectral statistics (LSS) of sample covariance matrices is another important topic in statistics and random matrix theory. The most influential work is [9]. They showed that the LSS of sample covariance matrices converge to normal distribution under some moment restrictions. Further refinements were carried out under different relaxed settings. [25] improved Theorem 1.1 in [9] by removing the constraint on the fourth moment of the underlying random variables. [24] showed the CLT of LSS for noncentered sample covariance matrices and discussed the difference between the centered and noncentered sample covariance matrices. [34] provided similar results for centralized and noncentralized sample covariance matrices in a unified framework. Furthermore [23] also provided CLT in terms of vanishing Lévy–Prohorov distance between the LSS distribution and a Gaussian probability distribution.

However, even though a lot of effort has been devoted to these two topics separately the relationship between the extreme eigenvalues and linear spectral statistics has not been well understood. [10] obtained the joint normal distribution of the largest eigenvalue and LSS for a spiked Wigner matrix. They showed that the asymptotic joint distribution of the largest eigenvalue and LSS converges to a bivariate normal distribution with the covariance dependent on the third moment of entries of the Wigner matrix. In this case, the spiked eigenvalues and LSS are asymptotically independent if the third moment is zero. Recently, [21] established that the extreme eigenvalues and the trace of sample covariance matrices are jointly asymptotically normal and independent for a block diagonal population covariance matrix.

This paper focuses on more general spiked covariance matrices instead of block diagonal population covariance matrices. Specifically speaking, we consider a population covariance matrix

$$\boldsymbol{\Sigma} = \mathbf{V}\begin{pmatrix} \boldsymbol{\Lambda}_S & 0 \\ 0 & \boldsymbol{\Lambda}_P \end{pmatrix}\mathbf{V}^{\mathsf{T}}, \tag{1.2}$$

where $\mathbf{V}$ is an orthogonal matrix, $\boldsymbol{\Lambda}_S$ is a diagonal matrix consisting of the bounded and descending spiked eigenvalues, and $\boldsymbol{\Lambda}_P$ is the diagonal matrix of nonspiked eigenvalues.

Our main contributions are summarized as follows. For the first time we establish CLT for the leading spiked eigenvalues of the sample covariance matrices with the general spiked covariance matrices $\boldsymbol{\Sigma}$ in (1.2). We need neither the block diagonal structure unlike [4, 5] nor the maximum absolute value of the eigenvector of the corresponding spikes tending to zero nor requiring the match of the fourth moment with the standard Gaussian distribution (i.e., the fourth moment is 3) unlike [16]. We also show that the extreme eigenvalues and LSS of large sample covariance matrices are asymptotically independent. Moreover consistent estimators of the L_4 norm of population eigenvectors associated with the leading sample spikes are proposed.

The remaining sections are organized as follows. Section 2 presents the main results about the asymptotic distribution of the largest sample spikes, the asymptotic independence between the largest sample spikes and the linear spectral statistics and the estimator of the population eigenvectors corresponding the largest spikes. We also explore an application of our main results in the two sample hypothesis testing about covariances in Section 2. The simulation is reported in Section 3.

Throughout the paper, we say that an event Ω_n holds with high probability if $P(\Omega_n) \geq 1 - O(n^{-l})$ for some large constant $l > 0$. We use $I(\mathcal{A})$ to denote an indicator function of an event $\mathcal{A}$. The intersection of events $\mathcal{A}$ and $\mathcal{B}$ is denoted by $\mathcal{A} \cap \mathcal{B}$, or abbreviated by $\mathcal{A}\mathcal{B}$. The spectral norm of a matrix M is denoted by $\|M\|$.

2. The main results. Consider the data matrix $\mathbf{\Gamma X}$, where $\mathbf{\Gamma}$ is a $p \times p$ deterministic matrix with $\mathbf{\Gamma}\mathbf{\Gamma}^\top = \mathbf{\Sigma}$ and $\mathbf{X} = (x_{ij})$ is a $p \times n$ random matrix with entries $x_{ij} = n^{-1/2} q_{ij}$ where q_{ij} are independent random variables satisfying Assumption 1 below. The sample covariance matrix has the form

$$\mathbf{S}_n = \mathbf{\Gamma}\mathbf{X}\mathbf{X}^\top\mathbf{\Gamma}^\top.$$

Order the eigenvalues of $\mathbf{S}_n$ as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. Denote the singular value decomposition of $\mathbf{\Gamma}$ by

$$\mathbf{\Gamma} = \mathbf{V} \begin{pmatrix} \mathbf{\Lambda}_S^{1/2} & 0 \\ 0 & \mathbf{\Lambda}_P^{1/2} \end{pmatrix} \mathbf{U}^\top \tag{2.1}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices, $\mathbf{\Lambda}_S$ is a diagonal matrix consisting of the spiked eigenvalues in descending order and $\mathbf{\Lambda}_P$ is the diagonal matrix of the nonspiked eigenvalues. To be more specific, denote the eigenvalues of the spiked part $\mathbf{\Lambda}_S$ as $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_K$, and eigenvalues of the nonspiked part as $\alpha_{K+1} \geq \alpha_{K+2} \cdots \geq \alpha_p$. Partition $\mathbf{U}$ as $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2)$, where $\mathbf{U}_1$ is a $p \times K$ submatrix of $\mathbf{U}$. Let $\mathbf{u}_i = (u_{i1}, \ldots, u_{ip})^\top$ be the ith column of $\mathbf{U}_1$. Define

$$\mathbf{\Sigma}_{1P} = \mathbf{U}_2 \mathbf{\Lambda}_P \mathbf{U}_2^\top. \tag{2.2}$$

2.1. *Limiting laws for spiked eigenvalues.* We first specify the assumptions for establishing CLTs of the leading sample spiked eigenvalues.

ASSUMPTION 1. The double array $\{q_{ij} : i = 1, \ldots, p, j = 1, \ldots, n\}$ consists of independent and identically distributed random variables, with $Eq_{11} = 0$, $E|q_{11}|^2 = 1$ and $E|q_{11}|^4 = \gamma_4$.

ASSUMPTION 2. $p/n = c_n \to c \in (0, \infty)$ as $n \to \infty$.

ASSUMPTION 3. The $p \times p$ matrix $\mathbf{\Sigma} = \mathbf{\Gamma}\mathbf{\Gamma}^\top$ has a bounded spectral norm. Furthermore, denote the empirical spectral distribution (ESD) of $\mathbf{\Sigma}$ by H_n, which tends to a nonrandom limiting distribution H as $p \to \infty$.

In Assumption 1, the distribution of q_{11} is fixed and not allowed to change with n. Actually the distribution of $\{q_{ij}\}$ can be depend on n. One may impose some Lindeberg-type conditions and some additional steps need to be taken to conclude the results in this paper. We will not pursue this in this work.

For the next assumption, we denote by Γ_μ the support for any measure μ on $\mathbb{R}$. For $\alpha \notin \Gamma_H$ and $\alpha \neq 0$, define

$$\psi(\alpha) := \alpha + c\alpha \int \frac{t}{\alpha - t} \, dH(t). \tag{2.3}$$

TABLE 1
Parameters in (2.1) and Assumptions 1–4 classified based on whether they are allowed to change with n

Changing with n	Parameters
Yes	$c_n, \alpha_{K+1}, \dots, \alpha_p, H_n, u_{ij}$
No	$q_{ij}, c, K, \alpha_1, \dots, \alpha_K, H$

Its derivative is

$$\psi'(\alpha) = 1 - c\int \frac{t^2}{(\alpha - t)^2}\, dH(t). \tag{2.4}$$

Define $\psi_n(\alpha)$ from (2.3) with H, c replaced by H_n, c_n.

ASSUMPTION 4. Let K be a fixed integer. Suppose that the population covariance matrix $\mathbf{\Sigma}$ has K fixed spiked eigenvalues: $\alpha_1 > \cdots > \alpha_K$ not changing with n, lying outside the support of H and satisfying $\psi'(\alpha_k) > 0$ for $1 \le k \le K$.

[5] provided a complete characterization of sample spikes according to the sign of $\psi'(\alpha)$. If $\psi'(\alpha) > 0$ then the corresponding sample spiked eigenvalues have limits outside the support of $F^{c,H}$, the limit of the empirical spectral distribution of $\mathbf{S}_n$. They called them distant spikes in this case. Here we need to clarify that although [5] assumed that the population covariance matrices are block diagonal, this assumption is not essential and can be removed. This is because their method of deriving almost sure convergence relies on their Propositions 3.1 and 3.2 and these two results regarding the exact separation first appeared in [7, 8] without a block diagonal structure of the population covariance matrices. We highlight here that the number of nonspiked eigenvalues is $p - K$ thus these nonspiked eigenvalues may change with n. We only need to require that the ESD of the nonspiked eigenvalues tends to H which is implied by $H_n \to H$. For clarification purposes, we add Table 1 to summarize whether the parameters introduced above in (2.1) and Assumptions 1–4 are allowed to change with n. In practice, the number of spiked eigenvalues K is typically unknown and needed to be estimated from data. When $\mathbf{\Sigma}$ has a block diagonal structure, [21] developed test statistics that can be used to determine the number of spiked eigenvalues. [14] developed tests to determine K for divergent spiked eigenvalue models. For a general spiked covariance model, determining K is still a challenging and interesting problem, which will be pursued in a future paper.

We will show that the sample spiked eigenvalues λ_i ($i = 1, \dots, K$) of $\mathbf{S}_n$ are associated with a random quadratic form given by the following equation (see the details given in the proof of Theorem 2.2):

$$\det\{\mathbf{\Lambda}_S^{-1} - \mathbf{U}_1\mathbf{X}(\lambda_i\mathbf{I} - \mathbf{X}^\top\mathbf{\Sigma}_{1P}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{U}_1^\top\} = 0. \tag{2.5}$$

Thus, our results rely on a new technique tool, a CLT for a type of random quadratic forms. The result in Theorem 2.1 is crucial to removing the block diagonal structure of the population covariance matrices (hence the proof of Theorems 2.2 and 2.3 below). It can be of independent interest.

THEOREM 2.1. *Suppose that Assumptions 1–3 hold. Moreover, suppose that the non-random orthogonal unit vectors* $\mathbf{w}_1$ *and* $\mathbf{w}_2$ *satisfy* $\mathbf{w}_1^\top\mathbf{U}_2 = \mathbf{w}_2^\top\mathbf{U}_2 = 0$ *and* $\mathbf{w}_1^\top\mathbf{w}_2 = 0$, *and* α *satisfies* $\psi'(\alpha) > 0$. *Then*

$$\frac{\sqrt{n}}{\tilde{\sigma}_1}\left(\mathbf{w}_1^\top\mathbf{X}\left(\mathbf{I} - \frac{1}{\psi_n(\alpha)}\mathbf{X}^\top\mathbf{\Sigma}_{1P}\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{w}_1 - \frac{\psi_n(\alpha)}{\alpha}\right) \xrightarrow{D} N(0, 1) \tag{2.6}$$

and

$$\frac{\sqrt{n}}{\tilde{\sigma}_{12}}\mathbf{w}_1^\top \mathbf{X}\left(\mathbf{I}-\frac{1}{\psi_n(\alpha)}\mathbf{X}^\top \mathbf{\Sigma}_{1P}\mathbf{X}\right)^{-1}\mathbf{X}^\top \mathbf{w}_2 \xrightarrow{D} N(0,1), \tag{2.7}$$

where $\tilde{\sigma}_1^2:=\psi^2(\alpha)\{(\gamma_4-3)\sum_{i=1}^p w_{1i}^4+2/\psi'(\alpha)\}/\alpha^2$, $\tilde{\sigma}_{12}^2:=\{(\gamma_4-3)\psi^2(\alpha)\}\sum_{i=1}^p w_{1i}^2 w_{2i}^2/\alpha^2+\psi^2(\alpha)/[\alpha^2\psi'(\alpha)]$ *and* w_{ij} *is the* j*th element of* $\mathbf{w}_i$, $i=1,2$.

We are ready to provide the central limit theorem for the sample spiked eigenvalues. We consider the case when the eigenvalues of $\mathbf{\Lambda}_S$ are all simple first. The following assumption provides the asymptotic variances and covariances for the spiked eigenvalues.

ASSUMPTION 5. Assume that for $i=1,\dots,K$ the following limits exist:

$$\sigma_i^2 := \lim_{p\to\infty}(\gamma_4-3)\frac{\alpha_i^2\{\psi'(\alpha_i)\}^2}{\psi^2(\alpha_i)}\sum_{j=1}^p u_{ij}^4+2\frac{\alpha_i^2\psi'(\alpha_i)}{\psi^2(\alpha_i)} \quad \text{and}$$

$$\sigma_{ij} := \lim_{p\to\infty}(\gamma_4-3)\frac{\alpha_i\alpha_j\psi'(\alpha_i)\psi'(\alpha_j)}{\psi(\alpha_i)\psi(\alpha_j)}\sum_{k=1}^p u_{ik}^2 u_{jk}^2.$$

From Hölder's inequality, we have $\gamma_4\ge 1$. This together with two simple facts $\sum u_{ij}^4\le 1$ and $\psi'(\alpha_i)<1$ implies that $\sigma_i^2\in(0,\infty)$.

THEOREM 2.2. *Let* $\theta_i=\psi_n(\alpha_i)$, $i=1,\dots,K$, *and denote*

$$\Lambda_K=\left(\sqrt{n}\frac{\lambda_1-\theta_1}{\theta_1},\dots,\sqrt{n}\frac{\lambda_K-\theta_K}{\theta_K}\right). \tag{2.8}$$

Suppose that Assumptions 1–5 hold. Then

$$\Lambda_K \xrightarrow{D} N\big(0,\mathbf{\Sigma}^{(K)}\big), \tag{2.9}$$

where $\mathbf{\Sigma}^{(K)}=(\mathbf{\Sigma}_{ij}^{(K)})$ *with*

$$\mathbf{\Sigma}_{ij}^{(K)}=\begin{cases}\sigma_i^2 & i=j,\\ \sigma_{ij} & i\ne j.\end{cases}$$

REMARK 1. Since the convergence rate of $c_n\to c$ and $H_n\to H$ can be arbitrarily slow, $\theta_i=\psi_n(\alpha_i)$ is used in the CLT, rather than $\psi(\alpha_i)$, which is the almost sure limit of λ_i.

REMARK 2. If one cares about the asymptotic distribution for an individual sample spiked eigenvalue, Assumption 5 is not needed since $\sqrt{n}(\lambda_i-\theta_i)/\theta_i$ can be normalized by $[(\gamma_4-3)\alpha_i^2\{\psi'(\alpha_i)\}^2\sum_{j=1}^p u_{ij}^4+2\alpha_i^2\psi'(\alpha_i)]/\psi^2(\alpha_i)$. Moreover, Assumption 5 is not necessary if we restate the asymptotic convergence results using the Lévy–Prokhorov metric, as in (2.10) stated below. To be more concrete, recall Lévy–Prokhorov metric π which characterize the distance of two probability measure μ and ν in a metric space (M,d) defined by

$$\pi(\mu,\nu):=\inf\{\varepsilon>0 \mid \mu(A)\le\nu(A^\varepsilon)+\varepsilon \text{ and } \nu(A)\le\mu(A^\varepsilon)+\varepsilon \text{ for all } A\in\mathcal{B}(M)\},$$

where $A^\epsilon=\{x\in M\mid \exists y\in A, d(x,y)<\epsilon\}$. Let $\mathbf{\Sigma}_p^{(K)}=((\mathbf{\Sigma}_p^{(K)})_{ij})$ be $K\times K$ matrices with

$$(\mathbf{\Sigma}_p^{(K)})_{ij}:=\begin{cases}\sigma_{p,i}^2 & i=j,\\ \sigma_{p,ij} & i\ne j,\end{cases}$$

where

$$\sigma_{p,i}^2 := (\gamma_4 - 3)\frac{\alpha_i^2\{\psi'(\alpha_i)\}^2}{\psi^2(\alpha_i)}\sum_{j=1}^p u_{ij}^4 + 2\frac{\alpha_i^2\psi'(\alpha_i)}{\psi^2(\alpha_i)},$$

$$\sigma_{p,ij} := (\gamma_4 - 3)\frac{\alpha_i\alpha_j\psi'(\alpha_i)\psi'(\alpha_j)}{\psi(\alpha_i)\psi(\alpha_j)}\sum_{k=1}^p u_{ik}^2 u_{jk}^2.$$

Then

$$\lim_{n\to\infty} \pi\big(\Lambda_K, N\big(0, \mathbf{\Sigma}_p^{(K)}\big)\big) \to 0. \tag{2.10}$$

The proof of (2.10) is at the end of the proof of Theorem 2.2.

REMARK 3. Compared with earlier asymptotic results on spiked eigenvalues of sample covariance matrices obtained by [4, 5] and [21], we do not assume a block diagonal structure on population covariance matrices. Moreover, [4, 5] and [16] did not consider the joint distribution of the different leading sample spiked eigenvalues corresponding to the different population eigenvalues. Instead they considered the joint distribution of the different leading sample spiked eigenvalues corresponding to the same population eigenvalues.

REMARK 4. The bounded spiked eigenvalues setting in this paper is different from the divergent spiked eigenvalues setting considered in [19, 29, 32] and [14]. If we let the spiked eigenvalues $\alpha_i \to \infty$, the asymptotic variances σ_i^2 and covariance σ_{ij} in Theorem 2.2 converge, respectively, to the same asymptotic variances and covariance defined in Assumption 4 of [14]. This suggests that the asymptotic normality in Theorem 2.2 implies the asymptotic normality established in [14] intuitively.

We next consider the case when the multiplicity of the spiked eigenvalues of $\mathbf{\Lambda}_S$ are more than one.

ASSUMPTION 6. Suppose that the population covariance matrix $\mathbf{\Sigma}$ has K spiked eigenvalues: $\alpha_1 > \cdots > \alpha_{\mathcal{L}}$ with respective multiplicities $m_1, \dots, m_{\mathcal{L}}$, laying outside the support of H, and satisfying $\psi'(\alpha_k) > 0$ for $1 \le k \le \mathcal{L}$. Furthermore, we assume that the following limits exist for $i = 1, \dots, \mathcal{L}$:

$$\begin{aligned} g(r_i, k_1, l_1, k_2, l_2) = {} & \lim_{p\to\infty}(\gamma_4 - 3)\frac{\alpha_i^2\{\psi'(\alpha_i)\}^2}{\psi^2(\alpha_i)}\sum_{j=1}^p u_{r_i+k_1,j}u_{r_i+l_1,j}u_{r_i+k_2,j}u_{r_i+l_2,j} \\ & + \frac{\alpha_i^2\psi'(\alpha_i)}{\psi^2(\alpha_i)}\big\{\big(\mathbf{u}_{r_i+k_1}^\top\mathbf{u}_{r_i+k_2}\big)\big(\mathbf{u}_{r_i+l_1}^\top\mathbf{u}_{r_i+l_2}\big) \\ & + \big(\mathbf{u}_{r_i+k_1}^\top\mathbf{u}_{r_i+l_2}\big)\big(\mathbf{u}_{r_i+k_2}^\top\mathbf{u}_{r_i+l_1}\big)\big\}, \end{aligned} \tag{2.11}$$

where $r_i := \sum_{j=0}^{i-1} m_j$, $m_0 = 0$ and $1 \le k_1, l_1, k_2, l_2 \le m_i$.

THEOREM 2.3. *Suppose that Assumptions 1, 2, 3 and 6 hold. Then*

$$\left(\sqrt{n}\frac{\lambda_{r_i+1} - \theta_i}{\theta_i}, \dots, \sqrt{n}\frac{\lambda_{r_i+m_i} - \theta_i}{\theta_i}\right) \tag{2.12}$$

converges weakly to the joint distribution of the eigenvalues of $m_i \times m_i$ Gaussian random matrix $\mathbf{G}_i$ with $E\mathbf{G}_i = 0$ and covariance of $(\mathbf{G}_i)_{k_1,l_1}$ and $(\mathbf{G}_i)_{k_2,l_2}$ being $g(r_i, k_1, l_1, k_2, l_2)$ defined in (2.11).

REMARK 5. This result is similar to those in [4, 5], Theorem 3.1, and Corollary 3.1 in [16]. However, we neither need a block diagonal population covariance structure as in [4, 5] nor the maximum absolute value of the eigenvector of the corresponding spikes tending to zero (i.e., $\max_{1\le i\le K,1\le j\le K}|u_{ij}|\to 0$) nor requiring the match of the fourth moment with Gaussian distribution (i.e., $\gamma_4 = 3$) as in [16]. The assumption [D] about the population eigenvectors in [16] excludes all the diagonal population covariance matrices when $\max_{1\le i\le K,1\le j\le K}|u_{ij}|\to 0$. Under their assumption [D], we have

$$g(r_i, k_1, l_1, k_2, l_2) = \begin{cases} 2\alpha_i^2\psi'(\alpha_i)/\psi^2(\alpha_i) & k_1 = k_2 = l_1 = l_2, \\ \alpha_i^2\psi'(\alpha_i)/\psi^2(\alpha_i) & k_1 = k_2 \text{ and } l_1 = l_2 \text{ or } k_1 = l_2 \text{ and } l_1 = k_2, \\ 0 & \text{otherwise,} \end{cases} \tag{2.13}$$

which is consistent with theirs.

At the end of this section, we consider the limiting laws for the sample spiked eigenvalues of the centralized sample covariance matrices

$$\mathcal{S}_n = \mathbf{\Gamma}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\intercal\mathbf{\Gamma}^\intercal, \tag{2.14}$$

where $\bar{\mathbf{x}}$ is the sample mean of $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Let $\mathbf{1}$ be a column vector with all the entries being 1, and define $\Phi = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}$. We can also write

$$\mathcal{S}_n = \mathbf{\Gamma}\mathbf{X}\Phi\mathbf{X}^\intercal\mathbf{\Gamma}^\intercal.$$

We have the following corollary.

COROLLARY 2.4. *Let λ_i^c $(i = 1, \ldots, K)$ be the spiked eigenvalues of $\mathcal{S}_n$. Theorems 2.2 and 2.3 still hold by replacing λ_i with λ_i^c.*

2.2. *Asymptotic joint distribution of sample spiked eigenvalues and linear spectral statistics.* We now turn to the asymptotic joint distribution of sample spiked eigenvalues and linear spectral statistics of sample covariance matrices. Some notations are introduced first. Let $F^{\mathbf{S}_n}$ be the ESD of the sample covariance matrix $\mathbf{S}_n$. It is well known that $F^{\mathbf{S}_n}$ under some mild assumptions converges weakly to a nonrandom distribution $F^{c,H}$ with probability one, whose Stieltjes transform is the unique solution in $\mathbb{C}^+$ to the equation

$$m = \int \frac{1}{t(1 - c - czm) - z}\,dH(t) \tag{2.15}$$

for $z \in \mathbb{C}^+$ (see [3]). We also need the nonasymptotic version of $F^{c,H}$ whose Stieltjes transform solves the above equation by replacing c and H with c_n and H_n respectively, and we denote it by F^{c_n,H_n}. Define the conjugate matrix of $\mathbf{S}_n$ by $\underline{\mathbf{S}}_n := \mathbf{X}^\intercal\mathbf{\Gamma}^\intercal\mathbf{\Gamma}\mathbf{X}$ that shares the same nonzero eigenvalues of $\mathbf{S}_n$. The ESD of $\underline{\mathbf{S}}_n$ has an almost sure limit whose Stieltjes transform $\underline{m} \equiv \underline{m}(z) \in \mathbb{C}^+$ is the unique solution for any $z \in \mathbb{C}^+$ to the equation

$$z = -\frac{1}{\underline{m}} + c\int \frac{t}{1 + t\underline{m}}\,dH(t). \tag{2.16}$$

Now we introduce the linear spectral statistics of sample covariance matrices defined by

$$L_p(f) := \sum_{i=1}^{p} f(\lambda_i) - p\int f(x)\,dF^{c_n,H_n}(x), \tag{2.17}$$

where $f(x)$ is an analytic function on an open interval containing

$$\left[\liminf_n \alpha_p I_{(0,1)}(c)(1-\sqrt{c})^2, \limsup_n \alpha_1(1+\sqrt{c})^2\right]. \tag{2.18}$$

ASSUMPTION 7. Let $\mathbf{e}_i$ be the $p \times 1$ column vector with the ith element being 1 and others being 0. Suppose that

$$\frac{1}{p}\sum_{i=1}^{p} \mathbf{e}_i^\top \boldsymbol{\Gamma}^\top \left(\underline{m}(z_1)\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top + \mathbf{I}\right)^{-1} \boldsymbol{\Gamma}\mathbf{e}_i \mathbf{e}_i^\top \boldsymbol{\Gamma}^\top \left(\underline{m}(z_2)\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top + \mathbf{I}\right)^{-1} \boldsymbol{\Gamma}\mathbf{e}_i \to h_1(z_1, z_2) \tag{2.19}$$

and

$$\frac{1}{p}\sum_{i=1}^{p} \mathbf{e}_i^\top \boldsymbol{\Gamma}^\top \left(\underline{m}(z)\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top + \mathbf{I}\right)^{-2} \boldsymbol{\Gamma}\mathbf{e}_i \mathbf{e}_i^\top \boldsymbol{\Gamma}^\top \left(\underline{m}(z)\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top + \mathbf{I}\right)^{-1} \boldsymbol{\Gamma}\mathbf{e}_i \to h_2(z). \tag{2.20}$$

REMARK 6. Assumption 7 has been used to establish the CLT for LSS, see Theorem 1.4 in [25] for $\mathbf{S}_n$ and Theorem 1 in [24] for $\mathcal{S}_n$. Remark 1.4 in [25] also tells that such an assumption can be removed under our Assumptions 1–3 together with either $\gamma_4 = 3$ or $\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma}$ is diagonal. Actually, if $\gamma_4 = 3$, the term in Assumption 7 will not appear in the variance (covariance) and one may see [9]. And if $\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma}$ is diagonal,

$$h_2(z) = \int \frac{t^2\, dH(t)}{(\underline{m}(z)t+1)^3},$$

$$h_1(z_1, z_2) = \int \frac{t^2\, dH(t)}{(m(z_1)t+1)(m(z_2)t+1)}.$$

THEOREM 2.5. *Suppose that Assumptions 1–5 and 7 hold, then the K-dimensional random vector Λ_K and $L_p(f)$ are jointly asymptotically normal and asymptotically independent. The same conclusion also holds for the centralized sample covariance matrices $\mathcal{S}_n$.*

REMARK 7. Let $f_1, \dots, f_k$ be k ($k < \infty$) different functions analytic on an open interval containing the interval defined in (2.18). The above result can be generalized to the joint normal distribution of $\Lambda_K, L_p(f_1), \dots, L_p(f_k)$ with the spiked eigenvalues part and the LSS part still being asymptotically independent. For the results regarding to marginal distribution of LSS, one can refer to [25] and [24]. Notice that the LSS for $\mathcal{S}_n$ and $\mathbf{S}_n$ have the same asymptotic covariance but different asymptotic mean.

REMARK 8. Compared with Theorem 3.1 in [21], we have two advantages. Firstly, we don't need the block diagonal assumption on the population covariance matrices. Secondly, our LSS is not restricted to the trace of sample covariance matrices.

2.3. *Estimating the population eigenvectors associated with the spiked eigenvalues.* This section is to explore the estimation of the population spiked eigenvectors associated with the simple spiked eigenvalues $\alpha_1, \dots \alpha_K$ involved in (2.9). Although many studies of the spiked eigenvectors have been carried out, most of them have not provided consistent estimators for the population eigenvectors in terms of certain norm. For example, [26] established the almost sure limit of $\mathbf{u}_i^\top \hat{\mathbf{u}}_i$ and a CLT for $\hat{\mathbf{u}}_i$ for any $1 \le i \le K$ under the assumption that $\mathbf{X}$ is Gaussian and $\boldsymbol{\Gamma}$ is diagonal with the nonspiked covariance being identity. [15] further characterized the limit of $\mathbf{u}_i^\top \hat{\mathbf{u}}_i$ for a general spiked model. However, these results are not helpful for estimating the population eigenvectors in terms of certain norm. Our following theorem provides a consistent estimator of $\sum_{k=1}^{p} u_{ik}^4$ inspired by the results in [22], which considered an estimator of $\mathbf{s}^\top \mathbf{u}_i$ where $\mathbf{s}$ is any fixed vector with a bounded norm in R^p when the underlying random variables are continuous with finite eighth order moments.

THEOREM 2.6. *Suppose that the assumptions of Theorem 2.2 hold and* $\mathbf{\Gamma}$ *is symmetric, that is, the left orthogonal matrix* $\mathbf{V}$ *in* (2.1) *equals* $\mathbf{U}$. *Let* $\hat{\mathbf{u}}_i$ *be eigenvectors of* $\mathbf{S}_n$ *associated with eigenvalue* λ_i *and* $\hat{u}_{ik}$ *be the kth coordinate of* $\hat{\mathbf{u}}_i$. *For* $1 \le i \le K$, $\sum_{k=1}^p u_{ik}^4$ *is consistently estimated by* $\sum_{j=1}^p \{\sum_{k=1}^p \theta_i(k)\hat{u}_{kj}^2\}^2$, *where*

$$\theta_i(k) = \begin{cases} -\phi_i(k) & k \neq i, \\ 1 + \varrho_i(k) & k = i, \end{cases}$$

$$\phi_i(k) = \frac{\lambda_i}{\lambda_k - \lambda_i} - \frac{\nu_i}{\lambda_k - \nu_i}, \tag{2.21}$$

$$\varrho_i(k) = \sum_{j \neq i}^{p} \left(\frac{\lambda_j}{\lambda_k - \lambda_j} - \frac{\nu_j}{\lambda_k - \nu_j} \right),$$

and where $\nu_1 \ge \nu_2 \ge \cdots \ge \nu_p$ *are the real valued solutions to the equation in* x:

$$\frac{1}{p} \sum_{i=1}^{p} \frac{\lambda_i}{\lambda_i - x} = \frac{1}{c}. \tag{2.22}$$

When $c > 1$, *take* $\nu_n = \cdots = \nu_p = 0$. *In the expressions of* $\phi_i(k)$ *and* $\varrho_i(k)$, *we use the convention that any term of form* $\frac{0}{0}$ *is* 0. *The conclusion also holds if the sample eigenvalues and eigenvectors of* $\mathbf{S}_n$ *are replaced by those of* $\mathcal{S}_n$ *defined in* (2.14) *correspondingly.*

REMARK 9. Table 4 below shows that such an estimator of $\sum_{k=1}^p u_{ik}^4$ is quite accurate.

2.4. *Testing the equality of two spiked covariance matrices.* This subsection is to explore an application of our results. Consider the problem of testing the equality of two spiked covariance matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$. Let $\{\mathbf{y}_{1i} = \mathbf{\Sigma}_1^{1/2}\mathbf{q}_{1i}, i = 1, \dots, n_1\}$ be i.i.d. p variate random samples from the population F_1 with mean zero and covariance matrix $\mathbf{\Sigma}_1$, and $\{\mathbf{y}_{2i} = \mathbf{\Sigma}_2^{1/2}\mathbf{q}_{2i}, i = 1, \dots, n_2\}$ be i.i.d. p variate random samples from the population F_2 with mean zero and covariance matrix $\mathbf{\Sigma}_2$. Suppose F_1 and F_2 are independent. Several tests on the hypothesis:

$$H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 \quad \text{versus} \quad H_1 : \mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2 \tag{2.23}$$

have been proposed under high-dimensional settings. To name a few, [20] suggested a test based on an unbiased estimator for $\operatorname{tr}[(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2)^2]$. The test in [13] is motivated by studying the maximum of standardized differences between entries of two sample covariance matrices to deal with sparse alternatives. [30] proposed a weighted statistic that is powerful for dense or sparse alternatives.

Let $\mathbf{Y}_1 = (\mathbf{y}_{11}, \dots, \mathbf{y}_{1n_1})$ and $\mathbf{Y}_2 = (\mathbf{y}_{21}, \dots, \mathbf{y}_{2n_2})$. Denote $\mathbf{x}_{1i} = n_1^{-1/2}\mathbf{y}_{1i}$, $i = 1, \dots, n_1$ and $\mathbf{x}_{2i} = {n_2}^{-1/2}\mathbf{y}_{2i}$, $i = 1, \dots, n_2$. Let $\mathbf{X}_1 = (\mathbf{x}_{11}, \dots, \mathbf{x}_{1n_1})$ and $\mathbf{X}_2 = (\mathbf{x}_{21}, \dots, \mathbf{x}_{2n_2})$. Denote two sample covariance matrices by

$$\mathbf{S}_1 = \frac{1}{n_1}\mathbf{Y}_1\mathbf{Y}_1^\intercal = \mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{X}_1\mathbf{X}_1^\intercal\mathbf{\Sigma}_1^{\frac{1}{2}} \quad \text{and} \quad \mathbf{S}_2 = \frac{1}{n_2}\mathbf{Y}_2\mathbf{Y}_2^\intercal = \mathbf{\Sigma}_2^{\frac{1}{2}}\mathbf{X}_2\mathbf{X}_2^\intercal\mathbf{\Sigma}_2^{\frac{1}{2}}. \tag{2.24}$$

We also assume that the respective largest spike eigenvalues of $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are simple for simplicity. Denote the largest eigenvalues of $\mathbf{S}_1$ and $\mathbf{S}_2$ as $\lambda_1(\mathbf{S}_1)$ and $\lambda_1(\mathbf{S}_2)$, respectively. Denote the largest spiked eigenvalues of $\mathbf{\Sigma}_k$ by $\alpha_{k1}, k = 1, 2$, and the corresponding eigenvector by $\mathbf{u}_{1,k} = (u_{11,k}, \dots, u_{1p,k})^\intercal$, $k = 1, 2$. To ensure that α_{k1} are spiked eigenvalues, we assume that

$$\psi_k'(\alpha_{k1}) > 0, \tag{2.25}$$

where ψ_k is obtained from ψ in (2.3) with H replaced by the limit of ESD of $\mathbf{\Sigma}_k$. Let $\gamma_{4k}, k = 1, 2$ be the fourth moment of $\{q_{1ij}, j = 1, \dots, p, i = 1, \dots, n_1\}$ and $\{q_{2ij}, j = 1, \dots, p, i = 1, \dots, n_2\}$, respectively. To speed up our exploration, we assume $n_1 = n_2 = n$ in the rest of this Section. The general case is discussed in Section S.7 in the supplementary material [33]. A natural test statistic for (2.23) by using the largest eigenvalues and the linear spectral statistics is

$$\left\{\sqrt{n}\frac{\lambda_1(\mathbf{S}_1) - \lambda_1(\mathbf{S}_2)}{\sigma_{\mathrm{spi}}}\right\}^2 + \left\{\frac{\mathrm{tr}(\mathbf{S}_1) + \mathrm{tr}(\mathbf{S}_1^2) - \mathrm{tr}(\mathbf{S}_2) - \mathrm{tr}(\mathbf{S}_2^2)}{\sigma_{\mathrm{lin}}}\right\}^2, \tag{2.26}$$

where

$$\sigma_{\mathrm{spi}}^2 = \sigma_{\mathrm{spi1}}^2 + \sigma_{\mathrm{spi2}}^2, \qquad \sigma_{\mathrm{lin}}^2 = \sigma_{\mathrm{lin1}}^2 + \sigma_{\mathrm{lin2}}^2,$$

$$\sigma_{\mathrm{spi}k}^2 = (\gamma_{4k} - 3)\alpha_{k1}^2\big(\psi'(\alpha_{k1})\big)^2 \sum_{j=1}^{p} u_{1j,k}^4 + 2\alpha_{k1}^2 \psi'(\alpha_{k1}), \quad k = 1, 2$$

and

$$\begin{aligned} \sigma_{\mathrm{lin}k}^2 &= 8c_n r_{k4} + 16c_n^2 r_{k3} r_{k1} + 8c_n r_{k3} + 8c_n^3 r_{k2}(r_{k1})^2 \\ &\quad + 8c_n^2 r_{k2} r_{k1} + 4c_n^2 (r_{k2})^2 + 2c_n r_{k2} \\ &\quad + (\gamma_{4k} - 3)\{4c_n r_{k4} + 8c_n^2 r_{k3} r_{k1} + 4c_n r_{k3} + 4c_n^3 r_{k2}(r_{k1})^2 \\ &\quad + 4c_n^2 r_{k2} r_{k1} + c_n r_{k2}\}, \quad k = 1, 2 \end{aligned} \tag{2.27}$$

with $r_{km} = \mathrm{tr}(\mathbf{\Sigma}_k^m)/p$ and $c_n = p/n$. The expression (2.27) is obtained by calculating the contour integral in (1.20) in [25]. This statistic is modified further below.

The statistic in (2.26) is asymptotic χ_2^2 under the null hypothesis by Theorem 2.5. We next develop the estimators of unknown parameters α_{1k}, $\psi'(\alpha_{1k})$, $\sum_{j=1}^p u_{1j,k}^4$, γ_{4k} and r_{km} for practical implementation. For notational simplicity, the population index k is omitted and we aim to find estimators of α_1, $\psi'(\alpha_1)$, $\sum_{j=1}^p u_{1j}^4$, $r_m = \mathrm{tr}\,\mathbf{\Sigma}_1^m/p$ and γ_4 associated with the population F_1. The similar estimators are applicable to F_2 as well.

The estimator of $\sum_{j=1}^p u_{1j}^4$ is given in Theorem 2.6. For the estimation of α_1, we use the result in [2]. Note that

$$\underline{m}_n^*(z) := -\frac{1 - c_n}{z} + \frac{1}{n}\sum_{j\ge 2}\frac{1}{\lambda_j - z} \xrightarrow{\text{a.s.}} \underline{m}(z), \qquad -\frac{1}{\underline{m}_n^*(\lambda_1)} \xrightarrow{\text{a.s.}} \alpha_1.$$

Therefore, as proposed by [2], α_1 is estimated by

$$\left(\frac{1 - c_n}{\lambda_1} + \frac{1}{n}\sum_{j\ge 2}\frac{1}{\lambda_1 - \lambda_j}\right)^{-1}. \tag{2.28}$$

Consider an estimator of $\psi'(\alpha_1)$ now. Since $\psi(\cdot)$ is the inverse of the function $\alpha : x \mapsto -1/\underline{m}(x)$, we obtain

$$\psi'(\alpha_1) = \frac{1}{\alpha_1^2 \underline{m}'\{\psi(\alpha_1)\}}. \tag{2.29}$$

Thus, we can estimate $\underline{m}'\{\psi(\alpha_1)\}$ by taking $z = \lambda_1$ in the expression of $d\underline{m}_n^*(z)/dz$, which is

$$\frac{1 - c_n}{\lambda_1^2} + \frac{1}{n}\sum_{j\ge 2}\frac{1}{(\lambda_j - \lambda_1)^2}. \tag{2.30}$$

An estimator of $\psi'(\alpha_1)$ follows by replacing α_1 with (2.28) and $\underline{m}'\{\psi(\alpha)\}$ with (2.30) in (2.29).

Let $s_m = \operatorname{tr}(\mathbf{S}_1^m)/p$. According to Lemma 2.16 in [31] and Theorem 1.4 in [25], we have the following consistent estimators A_m for r_m, $m = 1, 2, 3, 4$:

$$A_1 = s_1, \qquad A_2 = s_2 - c_n(A_1)^2, \qquad A_3 = s_3 - 3c_n A_1 A_2 - c_n^2(A_1)^3,$$
$$A_4 = s_4 - 2c_n(A_2)^2 - 4c_n A_1 A_3 - 6c_n^2(A_1)^2 A_2 - c_n^3(A_1)^4. \tag{2.31}$$

To estimate γ_4, notice that

$$\mathfrak{M} := \frac{1}{p} E\big(\mathbf{y}_{11}^\top \mathbf{y}_{11} - \operatorname{tr}\boldsymbol{\Sigma}_1\big)^2 = \frac{\gamma_4 - 3}{p}\sum_{i=1}^{p}(\boldsymbol{\Sigma}_{1ii})^2 + 2r_2, \tag{2.32}$$

where $\boldsymbol{\Sigma}_{1ii}$ and $\mathbf{S}_{1ii}$ are, respectively, the ith diagonal entry of $\boldsymbol{\Sigma}_1$ and $\mathbf{S}_1$. Since r_2 can be estimated by A_2 above, we just need to find estimators of $\mathfrak{M}$ and $\sum_{i=1}^{p}(\boldsymbol{\Sigma}_{1ii})^2/p$. The following lemma specifies their consistent estimators.

LEMMA 2.7. *Under Assumptions 1 and 2, and assuming that $\boldsymbol{\Sigma}_1$ has bounded spectral norm, we have*

$$\frac{1}{p}\sum_{i=1}^{p}\mathbf{S}_{1ii}^2 - \frac{1}{p}\sum_{i=1}^{p}(\boldsymbol{\Sigma}_{1ii})^2 \xrightarrow{p} 0, \tag{2.33}$$

and

$$\frac{1}{pn}\sum_{i=1}^{n}\big(\mathbf{y}_{1i}^\top \mathbf{y}_{1i} - \operatorname{tr}\mathbf{S}_1\big)^2 - \mathfrak{M} \xrightarrow{p} 0, \tag{2.34}$$

where $\mathbf{y}_{1i}$ denotes the ith observation from the first population.

We assume that $\sum_{i=1}^{p}(\boldsymbol{\Sigma}_{1ii})^2/p$ does not converge to 0, which is a mild assumption for a population covariance matrix (otherwise the variances of the majority of the underlying random variables tend to zero). From (2.32) and Lemma 2.7, we propose a consistent estimator for γ_4 as follows

$$\hat{\gamma}_4 = \frac{n^{-1}\sum_{i=1}^{n}(\mathbf{y}_{1i}^\top \mathbf{y}_{1i})^2 - (1 - 2/n)(\operatorname{tr}\mathbf{S}_1)^2 - 2\operatorname{tr}\mathbf{S}_1^2}{\sum_{i=1}^{p}\mathbf{S}_{1ii}^2} + 3. \tag{2.35}$$

Through our simulations, we find that the largest sample spiked eigenvalue and the full linear spectral statistics have large correlations although they are asymptotic uncorrelated in theory. This is due to the fact that $\operatorname{cov}(\sqrt{n}\lambda_1/\psi(\alpha_1), \lambda_1) = O\{\psi(\alpha)/\sqrt{n}\}$ by Theorem 2.2, which is theoretically negligible. However, in practice, it may happen that $\psi(\alpha_1)$ is comparable to $\sqrt{n}$ (e.g., $\psi(8) = 8 + 7c/8$ for model 1 in the simulation part) which results in significant covariance. Therefore, we correct the statistic in (2.26) by removing the largest sample eigenvalue from the linear spectral statistics part. Actually, by using Slutsky's theorem and the fact that the single sample spiked eigenvalues converge to a constant, our proof of Theorem 2.5 also applies to the case when linear spectral statistics do not include the sample spiked eigenvalues.

It then suffices to recalculate the variance of LSS part without the largest sample spiked eigenvalue and estimate it. By taking contour z enclosing all the sample eigenvalues except the largest spiked one, and after analyzing the contour integral, we find that the more accurate variance is just to replace $r_m = \operatorname{tr}\boldsymbol{\Sigma}^m/p$ with $r_m - \alpha_1^m/p$, $m = 1, 2, 3, 4$ in (2.27), and denote it by $(\sigma_{\text{lin}}^{(1)})^2$ (for population associated with $\boldsymbol{\Sigma}_k$ denote it by $(\sigma_{\text{link}}^{(1)})^2$). The corresponding estimator for $r_m - \alpha_1^m/p$ is obtained by replacing $s_m = \operatorname{tr}\mathbf{S}_1^m/n$ with $s_m - \lambda_1^m/n$ in (2.31). Thus, we find an estimator for $(\sigma_{\text{lin}}^{(1)})^2 = (\sigma_{\text{lin1}}^{(1)})^2 + (\sigma_{\text{lin2}}^{(1)})^2$, and denote it by $(\widehat{\sigma_{\text{lin}}^{(1)}})^2 = (\widehat{\sigma_{\text{lin1}}^{(1)}})^2 + (\widehat{\sigma_{\text{lin2}}^{(1)}})^2$. Let

$$
\begin{aligned}
M_n &= \sqrt{n}\frac{\lambda_1(\mathbf{S}_1) - \lambda_1(\mathbf{S}_2)}{\widehat{\sigma_{\text{spi}}}};\\
L_n &= \frac{\sum_{i=1}^p f\{\lambda_i(\mathbf{S}_1)\} - f\{\lambda_i(\mathbf{S}_2)\}}{\widehat{\sigma_{\text{lin}}}}, \quad \text{where } f(x) = x + x^2;\\
L_n^{(1)} &= \frac{\sum_{i=2}^p f\{\lambda_i(\mathbf{S}_1)\} - f\{\lambda_i(\mathbf{S}_2)\}}{\widehat{\sigma_{\text{lin}}^{(1)}}}, \quad \text{where } f(x) = x + x^2;\\
T_n &= M_n^2 + \big(L_n^{(1)}\big)^2.
\end{aligned} \tag{2.36}
$$

We then propose the above statistic T_n for the hypothesis testing (2.23). As discussed before, under H_0 in (2.23), M_n, L_n, $L_n^{(1)}$ are all asymptotically $N(0,1)$. We summarize the asymptotic distribution of T_n under H_0 by the following theorem.

THEOREM 2.8. *Suppose that $\mathbf{S}_1$ and $\mathbf{S}_2$ defined in (2.24) satisfy Assumptions 1, 2 and (2.25). Moreover we assume that $\sum_{i=1}^p(\boldsymbol{\Sigma}_{kii})^2/p, k = 1, 2$ are bounded from below for sufficiently large p. Under the null hypothesis: $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, we have*

$$T_n \xrightarrow{D} \chi_2^2.$$

REMARK 10. The assumption that $\sum_{i=1}^p(\boldsymbol{\Sigma}_{kii})^2/p, k = 1, 2$ are bounded from below assures that the estimators of γ_{4k}, $k = 1, 2$ are consistent, see the discussion above (2.35). We address that Assumption 5 is not necessary here since the normalization of the variance has been made in M_n, see Remark 2. Similarly Assumption 7 is not required.

REMARK 11. One referee suggested a promising idea to improve the power of our test statistics by using $\operatorname{tr}(\mathbf{S}_1 - \mathbf{S}_2)^2$ to replace $L_n^{(1)}$ since $L_n^{(1)}$ might not be always nonzero if $\mathbf{S}_1 \neq \mathbf{S}_2$. Although Theorem 2.5 does not directly imply that $\operatorname{tr}(\mathbf{S}_1 - \mathbf{S}_2)^2$ is asymptotic independent with the spiked sample eigenvalues, we conjecture that the asymptotic independence holds for $\operatorname{tr}(\mathbf{S}_1 - \mathbf{S}_2)^2$ and the spiked eigenvalues.

Next, we analysis the power of our test. For $k = 1, 2$, denote

$$\mu_k = \frac{1}{n}\operatorname{tr}^2(\boldsymbol{\Sigma}_k) + \operatorname{tr}(\boldsymbol{\Sigma}_k) + \operatorname{tr}\big(\boldsymbol{\Sigma}_k^2\big). \tag{2.37}$$

THEOREM 2.9. *In addition to the assumptions in Theorem 2.8, if we let either $|\mu_1 - \mu_2| \to \infty$ or $\sqrt{n}|\psi_1(\alpha_{11}) - \psi_2(\alpha_{21})| \to \infty$ hold under H_1, then*

$$P(T_n > q_\alpha | H_1) \to 1 \quad \text{as } n \to \infty, \tag{2.38}$$

where q_α denotes the α-upper quantile of a χ_2^2 distribution, that is, if $X \sim \chi_2^2$, then $P(X > q_\alpha) = \alpha$.

In the above discussion, we assume the data matrix has zero mean. Corollary 2.4 and Theorem 2.5 allows us to construct a similar statistic to T_n for centralized sample covariance matrices, which meets well for real applications. We can also obtain consistent estimators for those parameters involved in the asymptotic variances for the normalized spiked eigenvalue and LSS, by replacing quantities associated to $\mathbf{S}_n$ with quantities associated to the centralized matrix $\mathcal{S}_n$ correspondingly.

3. Simulations and an empirical study. This section includes simulations to verify the performance of the earlier proposed statistics and the accuracy of the estimator of the population eigenvectors corresponding to the spikes. An empirical study further demonstrates the performance of our proposed statistic T_n.

3.1. *Simulations.* We introduce five covariance models to be used in simulations.

- Model 1: $\mathbf{\Sigma}^{(1)} = \mathrm{diag}(8, 1, \dots, 1)_{p\times p}$.
- Model 2: $\mathbf{\Sigma}^{(2)} = \mathrm{diag}(6, 2, \dots, 2, 1, \dots, 1)_{p\times p}$ where the number of 2 is 10.
- Model 3: $\mathbf{\Sigma}^{(3)} = \mathbf{O}_p \,\mathrm{diag}(12, d_2, \dots, d_p)_{p\times p}\mathbf{O}_p^{\intercal}$ where $d_i = 3 - 1.5(i-1)/p$, and
$$\mathbf{O}_p = \begin{bmatrix} \mathbf{O}_1 & 0 \\ 0 & \mathbf{I}_{p-3} \end{bmatrix},$$
where $\mathbf{O}_1$ is a 3×3 orthogonal matrix.
- Model 4: $\mathbf{\Sigma}^{(4)} = \mathbf{O}_p \,\mathrm{diag}(15, d_2, \dots, d_p)_{p\times p}\mathbf{O}_p^{\intercal}$ where $d_i = 3 - 2(i-1)/p$ and $\mathbf{O}_p$ is the same as Model 3.
- Model 5: $\mathbf{\Sigma}^{(5)} = \mathrm{diag}(12, 2, \dots, 2, 1, \dots, 1)_{p\times p}$ where the number of 2 is 10.

Assumption 4 holds for the largest eigenvalue of these models under our setting of p, n in simulations. For the models 1 and 2, $H = \delta_1$ where δ_a is the Dirac measure at the point a. According to Assumption 4, for those population eigenvalues satisfying $\alpha_i > 1 + \sqrt{c}$, the associated sample eigenvalues are spiked eigenvalues. We will set $p/n \le 3$. Thus, in model 1, the largest eigenvalue ($\alpha_1 = 8$) satisfies Assumption 4. In the model 2, the largest population eigenvalue ($\alpha_1 = 6$) satisfies Assumption 4, while those eigenvalues that equal 2 are spiked eigenvalue when $c \le 1$. Since we only include the largest one in the spiked part of our statistic T_n, it doesn't matter whether these eigenvalues are spikes or not. For model 3, we may regard H as a uniform distribution from 3/2 to 3, then

$$\psi(\alpha) = \alpha + \frac{2c\alpha}{3}\int_{3/2}^{3} \frac{t}{\alpha - t}\,dt = (1-c)\alpha - \frac{2c\alpha^2}{3}\ln\frac{\alpha-3}{\alpha-3/2}.$$

Thus,

$$\psi'(\alpha) = 1 - c - \frac{4c\alpha}{3}\ln\frac{\alpha-3}{\alpha-3/2} - \frac{c\alpha^2}{(\alpha-3)(\alpha-3/2)}.$$

The largest population eigenvalue $\alpha_1 = 12$ satisfies Assumption 4 when $1 - 0.057c > 0$. Similarly $\alpha_1 = 15$ satisfies Assumption 4 for the model 4. The model 5 is similar to model 2 and the largest eigenvalue satisfies Assumption 4.

We consider two types of distribution for entries of $\mathbf{X}_1$ and $\mathbf{X}_2$: standard normal distribution, and $t_{10}/\sqrt{(5/4)}$. We investigate the performance of T_n, and compare it with the tests in [20] and [13], respectively, denoted as Chen's test and Cai's test. The performance of M_n and L_n are also reported.

3.2. *Approximation accuracy.* In Tables 2 and 3, we report the empirical sizes of testing $H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}^{(i)}$ for $\mathbf{\Sigma}^{(i)}$ given by the above Models 1–5. The results listed in Table 2 are for standard normal distributed entries while Table 3 is for normalized t_{10} distributed entries. We run 500 simulation replications for each test of population covariance matrices. The nominal test size is 0.05. From the tables, we can see that the empirical sizes are around 0.05, which indicates that the χ_2^2 approximation is accurate. We would like to point out that although $\lambda_1 \xrightarrow{\text{a.s.}} \psi(\alpha_1)$ as n goes to infinity, the approximation is not accurate enough when $n = 100$. The estimating errors in (2.30) and (2.31) are slightly amplified if λ_1^2 is involved. This accounts for the slightly smaller size for statistics T_n, M_n and L_n in Tables 2 and 3.

TABLE 2

Empirical sizes for testing $H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}^{(i)}$ for data generated from Model i $(i = 1, 2, 3, 4, 5)$ with $N(0, 1)$ entries. The sample size is 100 for both samples

Model	p	40	60	80	100	120	150	240	300
$\mathbf{\Sigma}^{(1)}$	T_n	0.042	0.048	0.038	0.050	0.042	0.042	0.052	0.038
	M_n	0.044	0.044	0.054	0.042	0.042	0.042	0.042	0.038
	L_n	0.026	0.030	0.046	0.042	0.038	0.030	0.066	0.044
	Cai	0.036	0.048	0.030	0.058	0.044	0.040	0.044	0.046
	Chen	0.084	0.078	0.072	0.068	0.070	0.060	0.048	0.032
$\mathbf{\Sigma}^{(2)}$	T_n	0.036	0.030	0.040	0.052	0.026	0.063	0.042	0.048
	M_n	0.028	0.032	0.052	0.042	0.040	0.042	0.044	0.046
	L_n	0.030	0.030	0.042	0.042	0.036	0.054	0.044	0.044
	Cai	0.044	0.038	0.048	0.058	0.044	0.048	0.044	0.038
	Chen	0.056	0.050	0.066	0.056	0.046	0.053	0.054	0.056
$\mathbf{\Sigma}^{(3)}$	T_n	0.024	0.030	0.046	0.034	0.040	0.044	0.042	0.068
	M_n	0.038	0.036	0.044	0.072	0.034	0.034	0.038	0.066
	L_n	0.040	0.032	0.036	0.046	0.050	0.038	0.050	0.070
	Cai	0.040	0.046	0.048	0.036	0.032	0.054	0.048	0.028
	Chen	0.062	0.060	0.050	0.056	0.058	0.050	0.044	0.038
$\mathbf{\Sigma}^{(4)}$	T_n	0.056	0.044	0.048	0.034	0.048	0.044	0.030	0.030
	M_n	0.046	0.040	0.043	0.040	0.044	0.044	0.032	0.030
	L_n	0.034	0.040	0.038	0.032	0.038	0.052	0.048	0.036
	Cai	0.046	0.056	0.050	0.040	0.054	0.054	0.026	0.040
	Chen	0.072	0.064	0.086	0.066	0.058	0.058	0.030	0.042
$\mathbf{\Sigma}^{(5)}$	T_n	0.028	0.052	0.040	0.046	0.060	0.056	0.046	0.058
	M_n	0.042	0.028	0.046	0.036	0.038	0.042	0.030	0.052
	L_n	0.026	0.024	0.042	0.032	0.042	0.034	0.038	0.040
	Cai	0.046	0.054	0.054	0.044	0.046	0.038	0.028	0.042
	Chen	0.074	0.062	0.090	0.054	0.064	0.052	0.078	0.056

In Table 4, we record the performance of our estimator of $\sum_{i=1}^{p} u_{1i}^4$ for Model 4. The sample size is fixed to be 100, and for each dimension case, we run 500 replications and list the mean and variance. It can be seen that the estimator performs well.

3.3. *Power discussion.* We consider the power of tests for comparing three pairs of covariances $\mathbf{\Sigma}^{(1)}$ vs $\mathbf{\Sigma}^{(2)}$, $\mathbf{\Sigma}^{(3)}$ vs $\mathbf{\Sigma}^{(4)}$ and $\mathbf{\Sigma}^{(2)}$ vs $\mathbf{\Sigma}^{(5)}$. The empirical powers of the above three comparisons are, respectively, summarized in Tables 5, 6 and 7. In Tables 5–7, we find that T_n always outperforms Cai's test and Chen's test. In Table 5, all the three tests that is, T_n, Chen and Cai's tests have competitive powers. However, in Table 6, both Cai and Chen's tests lose powers while T_n has significant better powers and the powers increase as p increases. For the test comparing $\mathbf{\Sigma}^{(2)}$ and $\mathbf{\Sigma}^{(5)}$ in Table 7, Cai's test loses powers, while T_n and Chen's tests have satisfactory powers. The performance of T_n is more stable than Chen's test as p increases, and T_n outperforms Chen's test for large enough p. In fact, we can infer from (2.3) that the limit of difference of two sample spiked eigenvalues increases as p increases, so it is understandable that M_n has good powers for large p cases.

We observe T_n has good powers whenever the differences between the two covariances are introduced by either the nonspike eigenvalues (Tables 5 and 6) or the spike eigenvalues (Table 7). Specifically, for the tests comparing $\mathbf{\Sigma}^{(1)}$ and $\mathbf{\Sigma}^{(2)}$ (Table 5), the main differences between $\mathbf{\Sigma}^{(1)}$ and $\mathbf{\Sigma}^{(2)}$ are from nonspike eigenvalues. Thus, M_n has relatively low powers in this scenario, and the powers of T_n in Table 5 are inherited from the difference of LSS excluding the largest population eigenvalue, that is, the statistic $L_n^{(1)}$. This phenomenon can

TABLE 3
Empirical sizes for testing $H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}^{(i)}$ *for data generated from Model* i $(i = 1, 2, 3, 4, 5)$ *with* $t_{10}/\sqrt{5/4}$ *entries. The sample size is* 100 *for both samples*

Model	p	40	60	80	100	120	150	240	300
$\mathbf{\Sigma}^{(1)}$	T_n	0.052	0.034	0.054	0.038	0.034	0.042	0.030	0.040
	M_n	0.048	0.028	0.042	0.036	0.042	0.038	0.036	0.040
	L_n	0.034	0.026	0.042	0.036	0.042	0.046	0.06	0.056
	Cai	0.036	0.040	0.040	0.022	0.022	0.032	0.034	0.020
	Chen	0.098	0.102	0.112	0.094	0.092	0.054	0.048	0.050
$\mathbf{\Sigma}^{(2)}$	T_n	0.050	0.026	0.020	0.042	0.058	0.046	0.048	0.038
	M_n	0.048	0.028	0.026	0.034	0.042	0.030	0.034	0.038
	L_n	0.054	0.026	0.042	0.048	0.046	0.052	0.052	0.052
	Cai	0.036	0.036	0.034	0.032	0.046	0.020	0.034	0.026
	Chen	0.086	0.058	0.082	0.080	0.072	0.052	0.056	0.050
$\mathbf{\Sigma}^{(3)}$	T_n	0.038	0.030	0.034	0.046	0.042	0.030	0.038	0.048
	M_n	0.044	0.036	0.044	0.042	0.034	0.026	0.040	0.018
	L_n	0.028	0.038	0.036	0.040	0.050	0.042	0.044	0.046
	Cai	0.028	0.032	0.032	0.020	0.030	0.032	0.038	0.028
	Chen	0.084	0.066	0.072	0.054	0.066	0.050	0.038	0.040
$\mathbf{\Sigma}^{(4)}$	T_n	0.040	0.038	0.046	0.034	0.046	0.046	0.038	0.040
	M_n	0.042	0.034	0.034	0.054	0.030	0.036	0.038	0.040
	L_n	0.018	0.018	0.032	0.026	0.032	0.020	0.046	0.044
	Cai	0.038	0.042	0.040	0.028	0.022	0.024	0.032	0.018
	Chen	0.108	0.078	0.066	0.090	0.056	0.094	0.058	0.064
$\mathbf{\Sigma}^{(5)}$	T_n	0.044	0.042	0.052	0.060	0.030	0.060	0.034	0.042
	M_n	0.022	0.042	0.056	0.032	0.024	0.038	0.036	0.026
	L_n	0.022	0.034	0.028	0.030	0.018	0.038	0.034	0.038
	Cai	0.026	0.038	0.030	0.038	0.034	0.046	0.028	0.032
	Chen	0.100	0.118	0.104	0.086	0.084	0.104	0.070	0.048

be also seen by comparing the powers of L_n and $L_n^{(1)}$. Note that L_n does not have good powers because the largest eigenvalue in $\mathbf{\Sigma}^{(2)}$ is smaller than that of $\mathbf{\Sigma}^{(1)}$, which offsets the effect of $L_n^{(1)}$. In Table 6, the powers of T_n are also mainly contributed by $L_n^{(1)}$, but different from the results in Table 5, L_n and $L_n^{(1)}$ both have powers close to 1 for large p. However, in Table 7, M_n has significant powers but $L_n^{(1)}$ loses power because $\mathbf{\Sigma}^{(2)}$ and $\mathbf{\Sigma}^{(5)}$ shares the same eigenvalues except the large difference between their largest spiked population eigenvalues. In this scenario, the powers of T_n are mainly due to the contribution of M_n.

TABLE 4
Empirical mean and variance of the proposed estimators for $\sum_{i=1}^{p} u_{1i}^4$, *with true value* 0.5317 *in Model* 4. *The sample size is* 100 *and the simulation replication is* 500

Data entries	p	40	60	80	100	120	150	240	300
$N(0, 1)$	mean	0.5387	0.5417	0.5390	0.5404	0.5443	0.5430	0.5532	0.5489
	var	0.0029	0.0035	0.0044	0.0041	0.0044	0.0057	0.0073	0.0092
$t_{10}/\sqrt{5/4}$	mean	0.5378	0.5414	0.5394	0.5368	0.5425	0.5442	0.5441	0.5567
	var	0.0039	0.0042	0.0044	0.0048	0.0053	0.0055	0.0074	0.0110

TABLE 5
Empirical powers for testing $H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ *where* $\mathbf{\Sigma}_1 = \mathbf{\Sigma}^{(1)}$ *and* $\mathbf{\Sigma}_2 = \mathbf{\Sigma}^{(2)}$ *with two types of data entries*: $N(0, 1)$ *and* $t_{10}/\sqrt{5/4}$. *The sample size is* 100 *for both samples*

Data entries	p	40	60	80	100	120	150	240	300
$N(0, 1)$	T_n	1.000	1.000	1.000	0.994	0.992	0.978	0.894	0.822
	M_n	0.248	0.266	0.224	0.218	0.218	0.246	0.264	0.220
	L_n	0.194	0.188	0.232	0.256	0.286	0.230	0.274	0.270
	$L_n^{(1)}$	1.000	1.000	1.000	1.000	0.996	0.978	0.884	0.844
	Cai	0.796	0.668	0.584	0.530	0.434	0.372	0.260	0.194
	Chen	0.852	0.786	0.722	0.606	0.548	0.448	0.332	0.240
$t_{10}/\sqrt{5/4}$	T_n	1.000	0.998	0.996	0.980	0.942	0.898	0.728	0.600
	M_n	0.148	0.164	0.150	0.160	0.198	0.136	0.194	0.136
	L_n	0.142	0.176	0.182	0.204	0.200	0.204	0.200	0.216
	$L_n^{(1)}$	1.000	0.996	0.994	0.984	0.966	0.924	0.748	0.632
	Cai	0.448	0.326	0.236	0.190	0.162	0.132	0.066	0.062
	Chen	0.814	0.774	0.710	0.586	0.522	0.458	0.346	0.258

3.4. *An empirical study.* We use the gene expression data set from breast cancer study by [28] to illustrate our test. The data, available from "Bioconductor", consists of gene expression patterns of 200 tumors of patients who were not treated by systemic therapy after surgery. Patients were classified into three groups based on the tumor grade. In group 1, there are 29 patients with a well-differentiated tumor. In group 2, there are 136 patients with a moderately differentiated tumor. In group 3, there are 35 patients with a poor differentiated tumor. To understand the differences in gene interactions among these three patient groups, it is of interest to test the large dimensional covariance matrices come from these three groups.

The breast cancer data contains 22,283 features. We select those features that have coefficients of variation in the range (0.28, 1) and at least 30% of the patients have intensities above five. After this screening procedure, there are 723 features selected for analysis. To make the sample sizes comparable among three groups and check the empirical sizes using real data sets, we selected subsamples from groups 2 and 3. Specifically, we consider three data matrices: the group 1 denoted by $\mathbf{Y}_1$, the first 29 patients in group 2, denoted by $\mathbf{Y}_2$, and

TABLE 6
Empirical powers for testing $H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ *where* $\mathbf{\Sigma}_1 = \mathbf{\Sigma}^{(3)}$ *and* $\mathbf{\Sigma}_2 = \mathbf{\Sigma}^{(4)}$ *with two types of data entries*: $N(0, 1)$ *and* $t_{10}/\sqrt{5/4}$. *The sample size is* 100 *for both samples*

Data entries	p	40	60	80	100	120	150	240	300
$N(0, 1)$	T_n	0.792	0.938	0.976	0.994	1.000	1.000	1.000	1.000
	M_n	0.176	0.176	0.156	0.150	0.142	0.132	0.098	0.084
	L_n	0.040	0.062	0.178	0.318	0.506	0.762	0.994	1.000
	$L_n^{(1)}$	0.818	0.964	0.986	0.998	1.000	1.000	1.000	1.000
	Cai	0.062	0.050	0.092	0.074	0.068	0.076	0.070	0.052
	Chen	0.250	0.236	0.206	0.172	0.174	0.150	0.130	0.160
$t_{10}/\sqrt{5/4}$	T_n	0.636	0.818	0.930	0.954	0.982	0.998	1.000	1.000
	M_n	0.158	0.146	0.104	0.114	0.136	0.116	0.072	0.074
	L_n	0.024	0.054	0.144	0.266	0.374	0.632	0.986	0.996
	$L_n^{(1)}$	0.686	0.862	0.952	0.960	0.992	1.000	1.000	1.000
	Cai	0.032	0.058	0.026	0.052	0.040	0.030	0.032	0.036
	Chen	0.248	0.236	0.212	0.178	0.216	0.184	0.168	0.122

TABLE 7
Empirical powers for testing $H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ where $\mathbf{\Sigma}_1 = \mathbf{\Sigma}^{(2)}$ and $\mathbf{\Sigma}_2 = \mathbf{\Sigma}^{(5)}$ with two types of data entries: $N(0, 1)$ and $t_{10}/\sqrt{5/4}$. The sample size is 100 for both samples

Data entries	p	40	60	80	100	120	150	240	300
$N(0, 1)$	T_n	0.870	0.858	0.864	0.842	0.842	0.820	0.830	0.844
	M_n	0.910	0.940	0.946	0.926	0.908	0.888	0.900	0.906
	L_n	0.882	0.892	0.878	0.856	0.818	0.772	0.722	0.704
	$L_n^{(1)}$	0.046	0.048	0.046	0.046	0.058	0.056	0.036	0.062
	Cai	0.236	0.140	0.130	0.110	0.098	0.096	0.076	0.058
	Chen	0.908	0.872	0.878	0.854	0.778	0.730	0.630	0.596
$t_{10}/\sqrt{5/4}$	T_n	0.706	0.644	0.670	0.702	0.694	0.672	0.670	0.634
	M_n	0.822	0.748	0.790	0.786	0.804	0.786	0.782	0.770
	L_n	0.756	0.696	0.688	0.682	0.668	0.600	0.546	0.490
	$L_n^{(1)}$	0.054	0.046	0.050	0.040	0.050	0.054	0.044	0.042
	Cai	0.096	0.062	0.058	0.058	0.060	0.050	0.034	0.034
	Chen	0.838	0.784	0.790	0.768	0.778	0.704	0.634	0.588

the first 29 patients in group 3, denoted by $\mathbf{Y}_3$. These three data matrices are of the same size 723×29 which facilitate the analysis by our method. We plot the histogram of eigenvalues in Figure 1. We observe an obvious spiked eigenvalue for sample covariance matrices of $\mathbf{Y}_1$ and $\mathbf{Y}_3$. For the sample covariance matrix of $\mathbf{Y}_2$ the largest eigenvalue may also be a spiked eigenvalue. Thus, it is reasonable to apply our method on these data.

To detect the differences in gene interactions among three groups, we applied our proposed test to compare any two covariances among $\mathbf{\Sigma}_1$, $\mathbf{\Sigma}_2$ and $\mathbf{\Sigma}_3$ using data matrices $\mathbf{Y}_1$, $\mathbf{Y}_2$ and $\mathbf{Y}_3$, where $\mathbf{\Sigma}_i$ is the covariance among the 723 genes in the ith ($i = 1, 2, 3$) patient group. At 5% nominal level, our proposed tests reject all the three null hypotheses $H_{01} : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$, $H_{02} : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_3$ and $H_{03} : \mathbf{\Sigma}_2 = \mathbf{\Sigma}_3$. This indicates the significant differences among three covariance matrices. However, both Chen's and Cai's test fail to detect the difference in groups 1 and 2, and Cai's test cannot detect the difference between groups 2 and 3.

We split the data in group 2 to check the empirical size. Specifically, we use $\mathbf{Y}_4$, $\mathbf{Y}_5$ and $\mathbf{Y}_6$ to denote, respectively, the data matrices formed by the 30th–58th patients, the 59th–87th patients, and the 88th–116th patients in group 2. Then, we applied our proposed test statistics T_n to any pairs among $\mathbf{Y}_2$, $\mathbf{Y}_4$, $\mathbf{Y}_5$ and $\mathbf{Y}_6$, the p-values of these tests are all larger than 5%

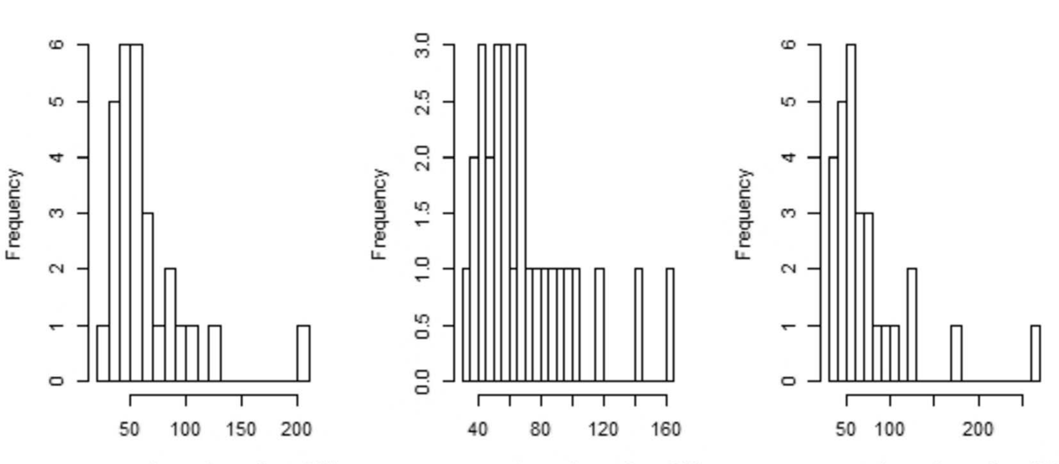

FIG. 1. *Histogram of nonzero eigenvalues of the covariance matrices of $\mathbf{Y}_1$, $\mathbf{Y}_2$ and $\mathbf{Y}_3$.*

and we cannot detect any difference among the covariance matrices. This is predictable since they all come from the second group. This also confirms that our test is able to control the type I error in this real data set.

4. Proof of Theorem 2.1. This section is to give the proof of Theorem 2.1. We begin with a list of results.

1. For general matrices **A**, **B** and column vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ that are conformable, we have the following matrix formulas:

$$\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{B}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{A}^{-1}, \tag{4.1}$$

$$\mathbf{A}(\mathbf{I} + \mathbf{B}\mathbf{A})^{-1} = (\mathbf{I} + \mathbf{A}\mathbf{B})^{-1}\mathbf{A}, \tag{4.2}$$

$$(\mathbf{A} + \boldsymbol{\alpha}\boldsymbol{\beta}^{\top})^{-1}\boldsymbol{\alpha} = \frac{\mathbf{A}^{-1}\boldsymbol{\alpha}}{1 + \boldsymbol{\beta}^{\top}\mathbf{A}^{-1}\boldsymbol{\alpha}}. \tag{4.3}$$

2. Let $X = (X_1, \dots, X_n)$, where X_i's are i.i.d. real random variables with mean zero and variance one. Let $\mathbf{A} = (a_{ij})_{n\times n}$ and $\mathbf{B} = (b_{ij})_{n\times n}$ be two real or complex matrices. Then we have an identity

$$\begin{aligned} &E(X^{\top}\mathbf{A}X - \operatorname{tr}\mathbf{A})(X^{\top}\mathbf{B}X - \operatorname{tr}\mathbf{B}) \\ &\quad = (E|X_1|^4 - 3)\sum_{i=1}^{n} a_{ii}b_{ii} + \operatorname{tr}\mathbf{A}\mathbf{B}^{\top} + \operatorname{tr}\mathbf{A}\mathbf{B}. \end{aligned} \tag{4.4}$$

LEMMA 4.1 (Theorem 35.12 of [12]). *Suppose that for each n, $Y_{n1}, Y_{n2}, \dots, Y_{nr_n}$ is a real martingale difference sequence with respect to an increasing σ-field $\{\mathcal{F}_{nj}\}$ having second moments. If as $n \to \infty$,* (i) $\sum_{j=1}^{r_n} E(Y_{nj}^2|\mathcal{F}_{n,j-1}) \xrightarrow{p} \sigma^2$ *where σ^2 is a positive constant, and for each $\epsilon > 0$,* (ii) $\sum_{j=1}^{r_n} E\{Y_{nj}^2 I(|Y_{nj}| > \epsilon)\} \to 0$, *then*

$$\sum_{j=1}^{r_n} Y_{nj} \xrightarrow{D} N(0, \sigma^2).$$

3. Suppose that entries of $\mathbf{x}$ is truncated at $\eta_n n^{1/4}$ and centralized, that is, $x_{ij} = \frac{1}{\sqrt{n}} q_{ij}$, where q_{ij} satisfying Assumption 1, are truncated at $\eta_n n^{1/4}$ and centralized. $\mathbf{M}$, $\mathbf{M}_1$ and $\mathbf{M}_2$ are $p \times p$ nonrandom matrices (or independent of $\mathbf{x}$). $\mathbf{w}$ is a $p \times 1$ nonrandom vector with a bounded spectral norm. We conclude the following simple results from Lemma 2.2 in [9]:

$$E\left|\mathbf{x}^{\top}\mathbf{M}\mathbf{x} - \frac{1}{n}\operatorname{tr}\mathbf{M}\right|^d \le C\|\mathbf{M}\|^d n^{-d/2}, \tag{4.5}$$

$$E\left|\mathbf{x}^{\top}\mathbf{M}_1\mathbf{w}\mathbf{w}^{\top}\mathbf{M}_2\mathbf{x} - \frac{1}{n}\mathbf{w}^{\top}\mathbf{M}_2\mathbf{M}_1\mathbf{w}\right|^d \le C\|\mathbf{M}_1\|^d\|\mathbf{M}_2\|^d\eta_n^{2d-4}n^{-d/2-1}, \tag{4.6}$$

$$E|\mathbf{x}^{\top}\mathbf{M}_1\mathbf{w}\mathbf{w}^{\top}\mathbf{M}_2\mathbf{x}|^d \le C\|\mathbf{M}_1\|^d\|\mathbf{M}_2\|^d\eta_n^{2d-4}n^{-d/2-1}. \tag{4.7}$$

PROOF OF THEOREM 2.1. We below only prove (2.6) and the proof of (2.7) is similar. The overall strategy of the proof is to decompose $\mathbf{w}_1^{\top}\mathbf{X}(\mathbf{I} - \mathbf{X}^{\top}\frac{\boldsymbol{\Sigma}_{1P}}{\psi_n(\alpha)}\mathbf{X})^{-1}\mathbf{X}^{\top}\mathbf{w}_1$ into summation of martingale differences and then apply Lemma 4.1. We assume that X has already been truncated at $\eta_n n^{1/4}$ and centralized according to the argument in S.7 in the Supplementary file.

CLT of the random part. Throughout the rest of the paper, let $\mathbf{x}_k$ be the kth ($k = 1, \dots, n$) column of $\mathbf{X}$, and $\mathbf{e}_k = (0, \dots, 0, 1, 0, \dots, 0)$ be an n-dimensional vector with the kth element being 1. We use C to denote constants which may change from line to line. Introduce notations

$$
\begin{aligned}
&\theta = \lim \psi_n(\alpha) = \psi(\alpha), \qquad \mathbf{X}_k = \mathbf{X} - \mathbf{x}_k \mathbf{e}_k^\intercal,\\
&\mathbf{X}_{jk} = \mathbf{X} - \mathbf{x}_k \mathbf{e}_k^\intercal - \mathbf{x}_j \mathbf{e}_j^\intercal, \qquad \tilde{\mathbf{\Sigma}}_1 = \frac{\mathbf{\Sigma}_{1P}}{\psi_n(\alpha)},\\
&\mathbf{A} = \mathbf{I}_n - \mathbf{X}^\intercal \tilde{\mathbf{\Sigma}}_1 \mathbf{X}, \qquad \mathbf{A}_k = \mathbf{I}_n - \mathbf{X}_k^\intercal \tilde{\mathbf{\Sigma}}_1 \mathbf{X}_k,\\
&\mathbf{D}_k = \mathbf{I}_p - \tilde{\mathbf{\Sigma}}_1 \mathbf{X}_k \mathbf{X}_k^\intercal, \qquad \mathbf{D} = \mathbf{I}_p - \tilde{\mathbf{\Sigma}}_1 \mathbf{X}\mathbf{X}^\intercal,\\
&\mathbf{D}_{jk} = \mathbf{I}_p - \tilde{\mathbf{\Sigma}}_1 \mathbf{X}_{jk} \mathbf{X}_{jk}^\intercal, \qquad \mathbf{B}_k = \mathbf{D}_k^{-1} \mathbf{w}_1 \mathbf{w}_1^\intercal (\mathbf{D}_k^\intercal)^{-1},\\
&\delta_k = \mathbf{x}_k^\intercal \mathbf{B}_k \mathbf{x}_k - \frac{1}{n} \operatorname{tr} \mathbf{B}_k,\\
&\alpha_k = \frac{1}{1 - \mathbf{x}_k^\intercal \tilde{\mathbf{\Sigma}}_1 (\mathbf{D}_k^\intercal)^{-1} \mathbf{x}_k}, \qquad \alpha_{jk} = \frac{1}{1 - \mathbf{x}_k^\intercal \tilde{\mathbf{\Sigma}}_1 (\mathbf{D}_{jk}^\intercal)^{-1} \mathbf{x}_k},\\
&\bar{\alpha}_k = \frac{1}{1 - \frac{1}{n} \operatorname{tr} \tilde{\mathbf{\Sigma}}_1 (\mathbf{D}_k^\intercal)^{-1}},\\
&\bar{\alpha}_{jk} = \frac{1}{1 - \frac{1}{n} \operatorname{tr} \tilde{\mathbf{\Sigma}}_1 (\mathbf{D}_{jk}^\intercal)^{-1}}, \qquad a_n = \frac{1}{1 - \frac{1}{n} E \operatorname{tr} \tilde{\mathbf{\Sigma}}_1 (\mathbf{D}_1^\intercal)^{-1}},\\
&a_{1n} = \frac{1}{1 - E \frac{1}{n} \operatorname{tr} \tilde{\mathbf{\Sigma}}_1 (\mathbf{D}_{12}^\intercal)^{-1}},\\
&\gamma_k = \mathbf{x}_k^\intercal \tilde{\mathbf{\Sigma}}_1 (\mathbf{D}_k^\intercal)^{-1} \mathbf{x}_k - \frac{1}{n} \operatorname{tr} \tilde{\mathbf{\Sigma}}_1 (\mathbf{D}_k^\intercal)^{-1},\\
&\gamma_{1k} = \mathbf{x}_k^\intercal \tilde{\mathbf{\Sigma}}_1 (\mathbf{D}_{1k}^\intercal)^{-1} \mathbf{x}_k - \frac{1}{n} \operatorname{tr} \tilde{\mathbf{\Sigma}}_1 (\mathbf{D}_{1k}^\intercal)^{-1}.
\end{aligned}
\tag{4.8}
$$

With the help of (4.2) and (4.3), it is not difficult to conclude the following two facts:

$$
\mathbf{e}_k^\intercal \mathbf{X}_k^\intercal = \mathbf{e}_k^\intercal \mathbf{A}_k^{-1} \mathbf{X}_k^\intercal = 0, \tag{4.9}
$$

$$
\mathbf{e}_k^\intercal \mathbf{A}^{-1} \mathbf{X}^\intercal = \mathbf{e}_k^\intercal \mathbf{X}^\intercal \mathbf{D}^{-1} = \mathbf{x}_k^\intercal \mathbf{D}_k^{-1} \alpha_k. \tag{4.10}
$$

Note that the quantities defined in (4.8) such as α_k, $\bar{\alpha}_k$ and a_n are not always bounded and the matrices such as $\mathbf{A}$, $\mathbf{D}_k$ are not always invertible. So we introduce events

$$
\begin{aligned}
&\mathcal{B}_1 = \{\|\tilde{\mathbf{\Sigma}}_1 \mathbf{X}\mathbf{X}^\intercal\| \le 1 - \epsilon\}, \qquad \mathcal{B}_{1k} = \{\|\tilde{\mathbf{\Sigma}}_1 \mathbf{X}_k \mathbf{X}_k^\intercal\| \le 1 - \epsilon\},\\
&\mathcal{B}_{1jk} = \{\|\tilde{\mathbf{\Sigma}}_1 \mathbf{X}_{jk} \mathbf{X}_{jk}^\intercal\| \le 1 - \epsilon\},\\
&\mathcal{B}_{2k} = \left\{\left|\mathbf{x}_k^\intercal \tilde{\mathbf{\Sigma}}_1 (\mathbf{D}_k^\intercal)^{-1} \mathbf{x}_k - \left(1 + \frac{1}{\theta \underline{m}(\theta)}\right)\right| < \epsilon\right\},\\
&\mathcal{B}_{2jk} = \left\{\left|\mathbf{x}_k^\intercal \tilde{\mathbf{\Sigma}}_1 (\mathbf{D}_{jk}^\intercal)^{-1} \mathbf{x}_k - \left(1 + \frac{1}{\theta \underline{m}(\theta)}\right)\right| < \epsilon\right\},\\
&\mathcal{B}_{3k} = \left\{\left|\frac{1}{n} \operatorname{tr} \tilde{\mathbf{\Sigma}}_1 (\mathbf{D}_k^\intercal)^{-1} - \left(1 + \frac{1}{\theta \underline{m}(\theta)}\right)\right| < \epsilon\right\},\\
&\mathcal{B}_{3jk} = \left\{\left|\frac{1}{n} \operatorname{tr} \tilde{\mathbf{\Sigma}}_1 (\mathbf{D}_{jk}^\intercal)^{-1} - \left(1 + \frac{1}{\theta \underline{m}(\theta)}\right)\right| < \epsilon\right\},
\end{aligned}
\tag{4.11}
$$

where ϵ is a small positive constant. Note that $\mathcal{B}_1 \subseteq \mathcal{B}_{1k} \subseteq \mathcal{B}_{1jk}$. Denote

$$\mathcal{B} = \mathcal{B}_1 \cap \left(\bigcap_{i=2,3} \bigcap_{k=1}^{n} \mathcal{B}_{ik} \right) \cap \left(\bigcap_{i=2,3} \bigcap_{1\le j\neq k\le n} \mathcal{B}_{ijk} \right). \tag{4.12}$$

Then we have following lemma and the proof is postponed to the Supplementary file.

LEMMA 4.2. *The event $\mathcal{B}$ holds with high probability* (*i.e.*, $P(\mathcal{B}) = 1 - n^{-l}$ *for any large constant* l).

This lemma ensures that it suffices to establish CLT of $\mathbf{w}_1^\intercal \mathbf{X}\mathbf{A}^{-1}\mathbf{X}^\intercal \mathbf{w}_1 I(\mathcal{B})$. When the event $\mathcal{B}$ holds the terms α_k, α_{jk}, $\bar{\alpha}_k$ and $\bar{\alpha}_{jk}$ defined in (4.8) are bounded. We remark here that the more accurate definition of a_n should be

$$a_n = \frac{1}{1 - \frac{1}{n} E \operatorname{tr} \tilde{\mathbf{\Sigma}}_1 (\mathbf{D}_1^\intercal)^{-1} I(\mathcal{B}_1)},$$

which is bounded for sufficient large n, see (S.7.37) in the proof of Lemma 4.2. The definition of a_n in (4.8) is just for notational simplicity. Another important fact of a_n is

$$\lim_{n\to\infty} a_n \to \frac{\psi(\alpha)}{\alpha}. \tag{4.13}$$

This is because we have $a_n \to -\theta \underline{m}(\theta)$ as $n \to \infty$, see (S.7.37). Recall that $\theta = \psi(\alpha)$. By the fact that ψ is the inverse function of $\alpha : x \mapsto -1/\underline{m}(x)$, we have $\underline{m}(\theta) = -1/\alpha$. The above comment for a_n also applies to a_{1n}, and the limit of a_{1n} is also $\theta\underline{m}(\theta)$. Let $E_0(\cdot)$ denote expectation, and $E_k(\cdot)$ denote the conditional expectation with respect to σ-field generated by $\mathbf{x}_1, \dots, \mathbf{x}_k$. We have

$$\begin{aligned}
&\sqrt{n}\{\mathbf{w}_1^\intercal \mathbf{X}\mathbf{A}^{-1}\mathbf{X}^\intercal \mathbf{w}_1 I(\mathcal{B}) - E\mathbf{w}_1^\intercal \mathbf{X}\mathbf{A}^{-1}\mathbf{X}^\intercal \mathbf{w}_1 I(\mathcal{B})\}\\
&\quad= \sqrt{n}\sum_{k=1}^{n}(E_k - E_{k-1})\{\mathbf{w}_1^\intercal \mathbf{X}\mathbf{A}^{-1}\mathbf{X}^\intercal \mathbf{w}_1 I(\mathcal{B})\}\\
&\quad= \sqrt{n}\sum_{k=1}^{n}(E_k - E_{k-1})\{\mathbf{w}_1^\intercal \mathbf{X}\mathbf{A}^{-1}\mathbf{X}^\intercal \mathbf{w}_1 I(\mathcal{B}) - \mathbf{w}_1^\intercal \mathbf{X}_k\mathbf{A}_k^{-1}\mathbf{X}_k^\intercal \mathbf{w}_1 I(\mathcal{B}_{1k})\}\\
&\quad= \sqrt{n}\sum_{k=1}^{n}(E_k - E_{k-1})\{\mathbf{w}_1^\intercal \mathbf{X}\mathbf{A}^{-1}\mathbf{X}^\intercal \mathbf{w}_1 I(\mathcal{B}) - \mathbf{w}_1^\intercal \mathbf{X}_k\mathbf{A}_k^{-1}\mathbf{X}_k^\intercal \mathbf{w}_1 I(\mathcal{B})\} + o_p(1)\\
&\quad= \sqrt{n}\sum_{k=1}^{n}(E_k - E_{k-1})\{\alpha_k \mathbf{x}_k^\intercal B_k \mathbf{x}_k I(\mathcal{B})\} + o_p(1),
\end{aligned} \tag{4.14}$$

where the last step uses the fact that by (4.9) and (4.10)

$$\begin{aligned}
&\mathbf{w}_1^\intercal \mathbf{X}\mathbf{A}^{-1}\mathbf{X}^\intercal \mathbf{w}_1 - \mathbf{w}_1^\intercal \mathbf{X}_k\mathbf{A}_k^{-1}\mathbf{X}_k^\intercal \mathbf{w}_1\\
&\quad= \mathbf{w}_1^\intercal(\mathbf{X} - \mathbf{X}_k)\mathbf{A}^{-1}\mathbf{X}^\intercal \mathbf{w}_1 + \mathbf{w}_1^\intercal \mathbf{X}_k(\mathbf{A}^{-1} - \mathbf{A}_k^{-1})\mathbf{X}^\intercal \mathbf{w}_1 + \mathbf{w}_1^\intercal \mathbf{X}_k \mathbf{A}_k^{-1}(\mathbf{X}^\intercal - \mathbf{X}_k^\intercal)\mathbf{w}_1\\
&\quad= \mathbf{w}_1^\intercal \mathbf{x}_k \mathbf{e}_k^\intercal \mathbf{A}^{-1}\mathbf{X}^\intercal \mathbf{w}_1 + \mathbf{w}_1^\intercal \mathbf{X}_k \mathbf{A}_k^{-1}(\mathbf{X}^\intercal \tilde{\mathbf{\Sigma}}_1 \mathbf{X} - \mathbf{X}_k^\intercal \tilde{\mathbf{\Sigma}}_1 \mathbf{X}_k)\mathbf{A}^{-1}\mathbf{X}^\intercal \mathbf{w}_1\\
&\quad= \mathbf{w}_1^\intercal \mathbf{x}_k \mathbf{x}_k^\intercal \mathbf{D}_k^{-1}\mathbf{w}_1 \alpha_k + \mathbf{w}_1^\intercal \mathbf{X}_k \mathbf{A}_k^{-1}(\mathbf{e}_k \mathbf{x}_k^\intercal \tilde{\mathbf{\Sigma}}_1 \mathbf{X} + \mathbf{X}_k^\intercal \tilde{\mathbf{\Sigma}}_1 \mathbf{x}_k \mathbf{e}_k^\intercal)\mathbf{A}^{-1}\mathbf{X}^\intercal \mathbf{w}_1\\
&\quad= \mathbf{w}_1^\intercal \mathbf{x}_k \mathbf{x}_k^\intercal \mathbf{D}_k^{-1}\mathbf{w}_1 \alpha_k + \mathbf{w}_1^\intercal \mathbf{X}_k \mathbf{X}_k^\intercal \tilde{\mathbf{\Sigma}}_1 (\mathbf{D}_k^\intercal)^{-1} \mathbf{x}_k \mathbf{e}_k^\intercal \mathbf{A}^{-1}\mathbf{X}^\intercal \mathbf{w}_1\\
&\quad= \mathbf{w}_1^\intercal (\mathbf{D}_k^\intercal)^{-1} \mathbf{x}_k \mathbf{x}_k^\intercal \mathbf{D}_k^{-1} \mathbf{w}_1 \alpha_k = \alpha_k \mathbf{x}_k^\intercal \mathbf{B}_k \mathbf{x}_k.
\end{aligned} \tag{4.15}$$

Note that $\alpha_k = \bar{\alpha}_k + \bar{\alpha}_k^2\gamma_k + \bar{\alpha}_k^2\gamma_k^2\alpha_k$. It follows that

$$(4.14) = \sqrt{n}\sum_{k=1}^{n}(E_k - E_{k-1})\Big[\big(\bar{\alpha}_k + \bar{\alpha}_k^2\gamma_k + \bar{\alpha}_k^2\gamma_k^2\alpha_k\big)\Big(\delta_k + \frac{1}{n}\operatorname{tr} B_k\Big)I(\mathcal{B})\Big]$$

$$= \sqrt{n}\sum_{k=1}^{n} E_k\Big[\Big(\bar{\alpha}_k\delta_k + \frac{1}{n}\bar{\alpha}_k^2\gamma_k \operatorname{tr} B_k\Big)I(\mathcal{B}_{1k}\mathcal{B}_{3k})\Big] \tag{4.16}$$

$$+ \sqrt{n}\sum_{k=1}^{n}(E_k - E_{k-1})\big[\big(\bar{\alpha}_k^2\gamma_k\delta_k + \bar{\alpha}_k^2\gamma_k^2\alpha_k x_k^\top B_k\mathbf{x}_k\big)I(\mathcal{B})\big] + o_p(1),$$

where in the second equality, we use $E_{k-1}\{\bar{\alpha}_k\delta_k I(\mathcal{B})\} = E_{k-1}\{\bar{\alpha}_k\delta_k I(\mathcal{B}_{1k}\mathcal{B}_{3k})\} + o_p(n^{-2}) = o_p(n^{-2})$, and similarly, $E_{k-1}\frac{1}{n}\bar{\alpha}_k^2\gamma_k \operatorname{tr} B_k I(\mathcal{B}) = E_{k-1}\frac{1}{n}\bar{\alpha}_k^2\gamma_k \operatorname{tr} B_k I(\mathcal{B}_{1k}\mathcal{B}_{3k}) + o_p(n^{-2}) = o_p(n^{-2})$. We below omit the indicator functions such as $I(\mathcal{B})$, $I(\mathcal{B}_{1k})$ for simplicity, but one should bear in mind that a suitable indicate function of events is needed whenever handling the inverses of random matrices.

Using the Burkholder inequality, (4.5) and (4.6), we have

$$E\left|\sqrt{n}\sum_{k=1}^{n}(E_k - E_{k-1})\gamma_k\delta_k\right|^2 \le Cn^2\big(E|\gamma_k|^4\big)^{\frac{1}{2}}\big(E|\delta_k|^4\big)^{\frac{1}{2}} = o(1). \tag{4.17}$$

By similar arguments, together with the fact that $\bar{\alpha}_k$ and α_k are bounded, we have

$$\sqrt{n}\sum_{k=1}^{n}(E_k - E_{k-1})\big(\bar{\alpha}_k^2\gamma_k^2\alpha_k\mathbf{x}_k^\top\mathbf{B}_k\mathbf{x}_k\big) = o_p(1), \qquad \sqrt{n}\sum_{k=1}^{n} E_k\Big(\frac{1}{n}\bar{\alpha}_k^2\gamma_k \operatorname{tr}\mathbf{B}_k\Big) = o_p(1).$$

Therefore, we only need to consider $\sqrt{n}\sum_{k=1}^{n} E_k\bar{\alpha}_k\delta_k = \sqrt{n}\sum_{k=1}^{n}(E_k - E_{k-1})(\bar{\alpha}_k\delta_k) = \sqrt{n}\sum_{k=1}^{n}(E_k - E_{k-1})[(\bar{\alpha}_k - a_n)\delta_k] + \sqrt{n}a_n\sum_{k=1}^{n} E_k\delta_k$. Similar to (4.17), it is easy to get

$$E\left|\sqrt{n}\sum_{k=1}^{n}(E_k - E_{k-1})\big[(\bar{\alpha}_k - a_n)\delta_k\big]\right|^2 = o(1).$$

Summarizing the above, we conclude that

$$\sqrt{n}\big(\mathbf{w}_1^\top\mathbf{X}\mathbf{A}^{-1}\mathbf{X}^\top\mathbf{w}_1 - E\mathbf{w}_1^\top\mathbf{X}\mathbf{A}^{-1}\mathbf{X}^\top\mathbf{w}_1\big) = \sqrt{n}\sum_{k=1}^{n} a_n E_k(\delta_k) + o_p(1). \tag{4.18}$$

Let $Y_k = E_k\delta_k = (E_k - E_{k-1})\delta_k$. By the fact that a_n is bounded, we obtain

$$\sum_{k=1}^{n} a_n^2 E\big(nY_k^2 I\big(|\sqrt{n}Y_k| \ge \epsilon\big)\big) \le \frac{C}{\epsilon^2}\sum_{k=1}^{n} E|\sqrt{n}Y_k|^4 \le \frac{Cn^2}{\epsilon^2}\sum_{k=1}^{n} E|\delta_k|^4 = o(1), \tag{4.19}$$

where in the last step we use (4.6) and $\eta_n \to 0$. By Lemma 4.1 it suffices to verify

$$\sum_{k=1}^{n} a_n^2 E_{k-1}\big(nY_k^2\big) \xrightarrow{p} \sigma^2. \tag{4.20}$$

It follows from (4.4) that

$$n\sum_{k=1}^{n} E_{k-1}\big(Y_k^2\big) = \frac{1}{n}(\gamma_4 - 3)\sum_{k=1}^{n}\sum_{i=1}^{p}\big(E_k(\mathbf{B}_k)_{ii}\big)^2 + \frac{2}{n}\sum_{k=1}^{n}\operatorname{tr}(E_k\mathbf{B}_k)^2. \tag{4.21}$$

It suffices to find the limits of the following two expressions:

$$\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{p}\left(\mathbf{e}_i^{\top}E_k\mathbf{D}_k^{-1}\mathbf{w}_1\mathbf{w}_1^{\top}(\mathbf{D}_k^{\top})^{-1}\mathbf{e}_i\right)^2, \tag{4.22}$$

$$\frac{2}{n}\sum_{k=1}^{n}\operatorname{tr}\left(E_k\mathbf{D}_k^{-1}\mathbf{w}_1\mathbf{w}_1^{\top}(\mathbf{D}_k^{\top})^{-1}\right)^2. \tag{4.23}$$

To find the limit of (4.22), it is equivalent to considering the limit of

$$\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{p}\left(\mathbf{e}_i^{\top}E\mathbf{D}_k^{-1}\mathbf{w}_1\mathbf{w}_1^{\top}E\mathbf{D}_k^{\top-1}\mathbf{e}_i\right)^2, \tag{4.24}$$

whose proof is given in the supplementary file. Let

$$\mathbf{T}=\mathbf{I}-E\frac{1}{n}\sum_{k=2}^{n}\alpha_{1k}\tilde{\mathbf{\Sigma}}_1. \tag{4.25}$$

Using (S.7.42) and the dominated convergence theorem, it is easy to verify that

$$\lim_{n\to\infty}E\frac{1}{n}\sum_{k=2}^{n}\alpha_{1k}\to-\theta\underline{m}(\theta)=\frac{\psi(\alpha)}{\alpha}.$$

Since α has a positive distance to the support of $\mathbf{\Sigma}_{1P}$, $\mathbf{T}$ is invertible for large n. Write

$$\begin{aligned}
&E\left(\mathbf{e}_i^{\top}\mathbf{D}_1^{-1}\mathbf{w}_1\right)-\mathbf{e}_i^{\top}\mathbf{T}^{-1}\mathbf{w}_1\\
&\quad=E\left[\mathbf{e}_i^{\top}\mathbf{T}^{-1}\left(\sum_{j\geq 2}\tilde{\mathbf{\Sigma}}_1\mathbf{x}_j\mathbf{x}_j^{\top}-E\frac{1}{n}\sum_{j\geq 2}\alpha_{1j}\tilde{\mathbf{\Sigma}}_1\right)\mathbf{D}_1^{-1}\mathbf{w}_1\right]\\
&\quad=\sum_{j\geq 2}E\left[\alpha_{1j}\mathbf{x}_j^{\top}\mathbf{D}_{1j}^{-1}\mathbf{w}_1\mathbf{e}_i^{\top}\mathbf{T}^{-1}\tilde{\mathbf{\Sigma}}_1\mathbf{x}_j-\frac{E\alpha_{1j}}{n}\mathbf{e}_i^{\top}\mathbf{T}^{-1}\tilde{\mathbf{\Sigma}}_1\mathbf{D}_1^{-1}\mathbf{w}_1\right]\\
&\quad=A_1+A_2+A_3,
\end{aligned} \tag{4.26}$$

where

$$\begin{aligned}
A_1&=\sum_{j\geq 2}E\left((\alpha_{1j}-\bar{\alpha}_{1j})\left(\mathbf{x}_j^{\top}\mathbf{D}_{1j}^{-1}\mathbf{w}_1\mathbf{e}_i^{\top}\mathbf{T}^{-1}\tilde{\mathbf{\Sigma}}_1\mathbf{x}_j-\frac{1}{n}\mathbf{e}_i^{\top}\mathbf{T}^{-1}\tilde{\mathbf{\Sigma}}_1\mathbf{D}_{1j}^{-1}\mathbf{w}_1\right)\right),\\
A_2&=\sum_{j\geq 2}E\left(\frac{1}{n}\alpha_{1j}\mathbf{e}_i^{\top}\mathbf{T}^{-1}\tilde{\mathbf{\Sigma}}_1(\mathbf{D}_{1j}^{-1}-\mathbf{D}_1^{-1})\mathbf{w}_1\right),\\
A_3&=\sum_{j\geq 2}E\left(\frac{1}{n}\alpha_{1j}\mathbf{e}_i^{\top}\mathbf{T}^{-1}\tilde{\mathbf{\Sigma}}_1(\mathbf{D}_1^{-1}-E\mathbf{D}_1^{-1})\mathbf{w}_1\right).
\end{aligned} \tag{4.27}$$

We prove $A_1=O(n^{-1})$ first. Using $\alpha_{1j}-\bar{\alpha}_{1j}=\bar{\alpha}_{1j}^2\gamma_{1j}+\bar{\alpha}_{1j}^2\gamma_{1j}^2\alpha_{1j}$, we can write $A_1=A_{11}+A_{12}$, where

$$\begin{aligned}
A_{11}&=\sum_{j\geq 2}E\bar{\alpha}_{1j}^2\gamma_{1j}\left(\mathbf{x}_j^{\top}\mathbf{D}_{1j}^{-1}\mathbf{w}_1\mathbf{e}_i^{\top}\mathbf{T}^{-1}\tilde{\mathbf{\Sigma}}_1\mathbf{x}_j-\frac{1}{n}\mathbf{e}_i^{\top}\mathbf{T}^{-1}\tilde{\mathbf{\Sigma}}_1\mathbf{D}_{1j}^{-1}\mathbf{w}_1\right),\\
A_{12}&=\sum_{j\geq 2}E\alpha_{1j}\bar{\alpha}_{1j}^2\gamma_{1j}^2\left(\mathbf{x}_j^{\top}\mathbf{D}_{1j}^{-1}\mathbf{w}_1\mathbf{e}_i^{\top}\mathbf{T}^{-1}\tilde{\mathbf{\Sigma}}_1\mathbf{x}_j-\frac{1}{n}\mathbf{e}_i^{\top}\mathbf{T}^{-1}\tilde{\mathbf{\Sigma}}_1\mathbf{D}_{1j}^{-1}\mathbf{w}_1\right).
\end{aligned} \tag{4.28}$$

Using (4.4), we obtain

$$
\begin{aligned}
&E\Big[\bar{\alpha}_{1j}^2\gamma_{1j}\Big(\mathbf{x}_j^\top\mathbf{D}_{1j}^{-1}\mathbf{w}_1\mathbf{e}_i^\top\mathbf{T}^{-1}\tilde{\boldsymbol{\Sigma}}_1\mathbf{x}_j-\frac{1}{n}\mathbf{e}_i^\top\mathbf{T}^{-1}\tilde{\boldsymbol{\Sigma}}_1\mathbf{D}_{1j}^{-1}\mathbf{w}_1\Big)\big|\mathbf{X}_j\Big] \\
&\quad\le\frac{C}{n^2}\sum_{k=1}^{p}(\tilde{\boldsymbol{\Sigma}}_1\mathbf{D}_{1j}^{-1})_{kk}(\mathbf{D}_{1j}^{-1}\mathbf{w}_1\mathbf{e}_i^\top\mathbf{T}^{-1}\tilde{\boldsymbol{\Sigma}}_1)_{kk}+C\frac{\operatorname{tr}(\tilde{\boldsymbol{\Sigma}}_1\mathbf{D}_{1j}^{-2}\mathbf{w}_1e_i^\top\mathbf{T}^{-1}\tilde{\boldsymbol{\Sigma}}_1)}{n^2} \\
&\qquad+\frac{\operatorname{tr}(\tilde{\boldsymbol{\Sigma}}_1\mathbf{D}_{1j}^{-1}\tilde{\boldsymbol{\Sigma}}_1^\top\mathbf{T}^{-1}\mathbf{e}_i\mathbf{w}_1^\top\mathbf{D}_{1j}^{\top-1})}{n^2}.
\end{aligned}\tag{4.29}
$$

The first summation is bounded by

$$
\begin{aligned}
&\Big(\sum_{k=1}^{p}|\mathbf{e}_k^\top\mathbf{D}_{1j}^{-1}\mathbf{w}_1|^2\Big)^{1/2}\Big(\sum_{k=1}^{p}|\mathbf{e}_i^\top\mathbf{T}^{-1}\tilde{\boldsymbol{\Sigma}}_1\mathbf{e}_k\mathbf{e}_k^\top\tilde{\boldsymbol{\Sigma}}_1\mathbf{D}_{1j}^{-1}\mathbf{e}_k|^2\Big)^{1/2} \\
&\quad\le\|\mathbf{D}_{1j}^{-1}\|^2\|\tilde{\boldsymbol{\Sigma}}_1\|^2\|\mathbf{T}^{-1}\|<C.
\end{aligned}\tag{4.30}
$$

Thus the first term in (4.29) is $O(n^{-2})$. By similar but easier arguments, the second and third term also have bounds of the same order, so that we can conclude that

$$
|A_{11}|=O(n^{-1}).\tag{4.31}
$$

For A_{12}, using (4.5) and (4.6), we have

$$
\begin{aligned}
|A_{12}|&\le\sum_{j\ge2}(E|\alpha_{1j}\bar{\alpha}_{1j}^2\gamma_{1j}^2|^2)^{1/2} \\
&\quad\times\Big(E\Big|\mathbf{x}_j^\top\mathbf{D}_{1j}^{-1}\mathbf{w}_1\mathbf{e}_i^\top\mathbf{T}^{-1}\tilde{\boldsymbol{\Sigma}}_1\mathbf{x}_j-\frac{1}{n}\mathbf{e}_i^\top\mathbf{T}^{-1}\tilde{\boldsymbol{\Sigma}}_1\mathbf{D}_{1j}^{-1}\mathbf{w}_1\Big|^2\Big)^{1/2} \\
&=O(n^{-1}).
\end{aligned}\tag{4.32}
$$

Thus,

$$
|A_1|=O(n^{-1}).\tag{4.33}
$$

Consider the terms A_2 and A_3 now in (4.27). It follows from (4.1), (4.6), (4.7) and the Burkholder inequality that

$$
\begin{aligned}
|A_2|&=\Big|\sum_{j\ge2}E\Big(\frac{1}{n}\alpha_{1j}^2\mathbf{e}_i^\top\mathbf{T}^{-1}\tilde{\boldsymbol{\Sigma}}_1\mathbf{D}_{1j}^{-1}\tilde{\boldsymbol{\Sigma}}_1\mathbf{x}_j\mathbf{x}_j^\top\mathbf{D}_{1j}^{-1}\mathbf{w}_1\Big)\Big|=O(n^{-1}), \\
|A_3|&=\Big|\sum_{j\ge2}E\Big[\frac{1}{n}(\alpha_{1j}-a_n)\mathbf{e}_i^\top\mathbf{T}^{-1}\tilde{\boldsymbol{\Sigma}}_1(\mathbf{D}_1^{-1}-E\mathbf{D}_1^{-1})\mathbf{w}_1\Big]\Big| \\
&=\Big|\sum_{j\ge2}E\Big[\frac{1}{n}(\alpha_{1j}-a_n)\mathbf{e}_i^\top\mathbf{T}^{-1}\tilde{\boldsymbol{\Sigma}}_1\Big(\sum_{k=2}^{n}(E_k-E_{k-1})(\mathbf{D}_1^{-1}-\mathbf{D}_{1k}^{-1})\Big)\mathbf{w}_1\Big]\Big| \\
&\le\frac{1}{n}\sum_{j\ge2}(E|\alpha_{1j}-a_n|^2)^{\frac12} \\
&\quad\times\Big(E\Big|\sum_{k=1}^{n}(E_k-E_{k-1})\mathbf{e}_i^\top\mathbf{T}^{-1}\tilde{\boldsymbol{\Sigma}}_1\mathbf{D}_{1k}^{-1}\tilde{\boldsymbol{\Sigma}}_1\mathbf{x}_k\mathbf{x}_k^\top\mathbf{D}_{1k}^{-1}\mathbf{w}_1\Big|^2\Big)^{\frac12} \\
&=O(n^{-1}).
\end{aligned}\tag{4.34}
$$

From (4.26), (4.33), (4.34) and $\mathbf{w}_1^\intercal \mathbf{U}_2 = 0$, we have

$$E(\mathbf{e}_i^\intercal \mathbf{D}_1^{-1}\mathbf{w}_1) = \mathbf{e}_i^\intercal \mathbf{T}^{-1}\mathbf{w}_1 + O(n^{-1}) \\ = \mathbf{e}_i^\intercal \left(\mathbf{I} + \mathbf{T}^{-1} E \frac{1}{n}\sum_{k=2}^{n} \alpha_{1k}\tilde{\Sigma}_1\right)\mathbf{w}_1 + O(n^{-1}) \\ = w_{1i} + O(n^{-1}). \tag{4.35}$$

Substituting these back into (4.24), we conclude that (4.22) asymptotically equals $(1 + o(1))\sum_{i=1}^{p} w_{1i}^4$. For (4.23), it has the similar form to equation (4.7) in [6]. Hence, rewrite (4.23) as

$$\frac{2}{n}\sum_{k=1}^{n} E_{k-1} \operatorname{tr} E_k(\mathbf{D}_k^{-1}\mathbf{w}_1\mathbf{w}_1^\intercal \mathbf{D}_k^{\intercal -1}) E_k(\mathbf{D}_k^{-1}\mathbf{w}_1\mathbf{w}_1^\intercal \mathbf{D}_k^{\intercal -1}) \\ = \frac{2}{n}\sum_{k=1}^{n} E_{k-1}(\mathbf{w}_1^\intercal \mathbf{D}_k^{\intercal -1}\breve{\mathbf{D}}_k^{-1}\mathbf{w}_1\mathbf{w}_1^\intercal \breve{\mathbf{D}}_k^{\intercal -1}\mathbf{D}_k^{-1}\mathbf{w}_1), \tag{4.36}$$

where $\breve{\mathbf{D}}_k^{-1}$ is defined similarly as $\mathbf{D}_k^{-1}$ by $(\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}, \breve{\mathbf{x}}_{k+1}, \ldots, \breve{\mathbf{x}}_n)$ and where $\breve{\mathbf{x}}_{k+1}, \ldots, \breve{\mathbf{x}}_n$ are i.i.d. copies of $\mathbf{x}_{k+1}, \ldots, \mathbf{x}_n$. Following the argument similar to (4.7)–(4.22) of their work, we can obtain

$$E_{k-1}(\mathbf{w}_1^\intercal \mathbf{D}_k^{\intercal -1}\breve{\mathbf{D}}_k^{-1}\mathbf{w}_1\mathbf{w}_1^\intercal \breve{\mathbf{D}}_k^{\intercal -1}\mathbf{D}_k^{-1}\mathbf{w}_1) \times \left[1 - \frac{k-1}{n} a_{1n}^2 \frac{1}{n}\operatorname{tr}\tilde{\boldsymbol{\Sigma}}_1\mathbf{T}^{-1}\tilde{\boldsymbol{\Sigma}}_1\mathbf{T}^{-1}\right] \\ = (\mathbf{w}_1^\intercal \mathbf{T}^{-2}\mathbf{w}_1)^2\left[1 + \frac{k-1}{n} a_{1n}^2 \frac{1}{n} E_{k-1}\operatorname{tr}\mathbf{D}_k^{-1}\tilde{\boldsymbol{\Sigma}}_1\breve{\mathbf{D}}_k^{-1}\tilde{\boldsymbol{\Sigma}}_1\right] + o_p(1). \tag{4.37}$$

Note that $\mathbf{w}_1^\intercal \mathbf{T}^{-2}\mathbf{w}_1 = 1$ as in (4.35). Combining (4.2) with (2.18) of [9] we have

$$E_{k-1}\operatorname{tr}\mathbf{D}_k^{-1}\tilde{\boldsymbol{\Sigma}}_1\breve{\mathbf{D}}_k^{-1}\tilde{\boldsymbol{\Sigma}}_1 = \frac{\operatorname{tr}\tilde{\boldsymbol{\Sigma}}_1\mathbf{T}^{-1}\tilde{\boldsymbol{\Sigma}}_1\mathbf{T}^{-1} + o_p(1)}{1 - \frac{k-1}{n^2}\theta^2(\underline{m}(\theta))^2 \operatorname{tr}\tilde{\boldsymbol{\Sigma}}_1\mathbf{T}^{-1}\tilde{\boldsymbol{\Sigma}}_1\mathbf{T}^{-1}}. \tag{4.38}$$

Recall that $a_{1n} \to -\theta\underline{m}(\theta)$, $\underline{m}(\theta) \to \underline{m}(\psi(\alpha)) = -\alpha^{-1}$, and $F^{\boldsymbol{\Sigma}_{1P}} \to H$. Hence

$$d := \lim \frac{a_{1n}^2}{n}\operatorname{tr}\tilde{\boldsymbol{\Sigma}}_1\mathbf{T}^{-1}\tilde{\boldsymbol{\Sigma}}_1\mathbf{T}^{-1} \\ = \underline{m}^2(\theta)\int \frac{ct^2}{(1 + t\underline{m}(\theta))^2}\,dH(t) \\ = \int \frac{ct^2}{(\alpha - t)^2}\,dH(t) = 1 - \psi'(\alpha). \tag{4.39}$$

By (4.37) and (4.39), and similar argument to (4.35), we obtain

$$(4.36) \to 2(\mathbf{w}_1^\intercal \mathbf{T}^{-2}\mathbf{w}_1)^2\left(\int_0^1 \frac{1}{1 - td}\,dt + \int_0^1 \frac{td}{(1 - td)^2}\,dt\right) = \frac{2}{1 - d} = \frac{2}{\psi'(\alpha)}. \tag{4.40}$$

Consequently, from (4.21)–(4.40), and $a_n \to \psi(\alpha)/\alpha$, we conclude that

$$\sum_{k=1}^{n} a_n^2 E_{k-1}(nY_k^2) \to \frac{\psi^2(\alpha)}{\alpha^2}\left[(\gamma_4 - 3)\sum_{i=1}^{p} w_{1i}^4 + \frac{2}{\psi'(\alpha)}\right]. \tag{4.41}$$

Calculation of the mean. We can show that

$$\sqrt{n}\left(E\mathbf{w}_1^\intercal \mathbf{X}\mathbf{A}^{-1}\mathbf{X}^\intercal \mathbf{w}_1 - \frac{\psi_n(\alpha)}{\alpha}\right) \to 0. \tag{4.42}$$

Its proof is deferred to the Supplementary file. Consequently Theorem 2.1 is concluded. □

Acknowledgements. We are grateful to the Editor, the Associate Editor and two referees for their constructive comments, which helped us to improve the manuscript.

Funding. G. M. Pan was partially supported by MOE Tier 2 Grant 2018-T2-2-112 and by a MOE Tier 1 Grant RG133/18 at the Nanyang Technological University, Singapore.

S. Zheng (corresponding author) was partially supported by NSFC grant 12071066 and KLAS.

P.-S. Zhong was partially supported by an NSF grant DMS-1462156 and an NIH grant 1R21HG010073.

SUPPLEMENTARY MATERIAL

Supplementary file (DOI: 10.1214/22-AOS2183SUPP; .pdf). The file includes the proofs of (4.24), (4.42), Theorems 2.2–2.3, Corollary 2.4, Theorems 2.5–2.6, Lemma 2.7, Theorems 2.8–2.9 and Lemma 4.2, and an extension of the two sample tests in Section 2.4.

REFERENCES

[1] ANDERSON, T. W. (1962). An introduction to multivariate statistical analysis. Technical Report. Wiley, New York.

[2] BAI, Z. and DING, X. (2012). Estimation of spiked eigenvalues in spiked models. *Random Matrices Theory Appl.* **1** 1150011. MR2934717 https://doi.org/10.1142/S2010326311500110

[3] BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2567175 https://doi.org/10.1007/978-1-4419-0661-8

[4] BAI, Z. and YAO, J. (2008). Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 447–474. MR2451053 https://doi.org/10.1214/07-AIHP118

[5] BAI, Z. and YAO, J. (2012). On sample eigenvalues in a generalized spiked population model. *J. Multivariate Anal.* **106** 167–177. MR2887686 https://doi.org/10.1016/j.jmva.2011.10.009

[6] BAI, Z. D., MIAO, B. Q. and PAN, G. M. (2007). On asymptotics of eigenvectors of large sample covariance matrix. *Ann. Probab.* **35** 1532–1572. MR2330979 https://doi.org/10.1214/009117906000001079

[7] BAI, Z. D. and SILVERSTEIN, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26** 316–345. MR1617051 https://doi.org/10.1214/aop/1022855421

[8] BAI, Z. D. and SILVERSTEIN, J. W. (1999). Exact separation of eigenvalues of large-dimensional sample covariance matrices. *Ann. Probab.* **27** 1536–1555. MR1733159 https://doi.org/10.1214/aop/1022677458

[9] BAI, Z. D. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792 https://doi.org/10.1214/aop/1078415845

[10] BAIK, J., LEE, J. O. and WU, H. (2018). Ferromagnetic to paramagnetic transition in spherical spin glass. *J. Stat. Phys.* **173** 1484–1522. MR3878351 https://doi.org/10.1007/s10955-018-2150-6

[11] BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680 https://doi.org/10.1016/j.jmva.2005.08.003

[12] BILLINGSLEY, P. (1995). *Probability and Measure*. Wiley, New York.

[13] CAI, T., LIU, W. and XIA, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Amer. Statist. Assoc.* **108** 265–277. MR3174618 https://doi.org/10.1080/01621459.2012.758041

[14] CAI, T. T., HAN, X. and PAN, G. (2020). Limiting laws for divergent spiked eigenvalues and largest nonspiked eigenvalue of sample covariance matrices. *Ann. Statist.* **48** 1255–1280. MR4124322 https://doi.org/10.1214/18-AOS1798

[15] DING, X. (2017). Asymptotics of empirical eigen-structure for high dimensional sample covariance matrices of general form. arXiv preprint. Available at arXiv:1708.06296.

[16] JIANG, D. and BAI, Z. (2021). Generalized four moment theorem and an application to CLT for spiked eigenvalues of high-dimensional covariance matrices. *Bernoulli* **27** 274–294. MR4177370 https://doi.org/10.3150/20-BEJ1237

[17] JIANG, D. and BAI, Z. (2021). Partial generalized four moment theorem revisited. *Bernoulli* **27** 2337–2352. MR4303885 https://doi.org/10.3150/20-BEJ1310

[18] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544

[19] JUNG, S. and MARRON, J. S. (2009). PCA consistency in high dimension, low sample size context. *Ann. Statist.* **37** 4104–4130. MR2572454 https://doi.org/10.1214/09-AOS709

[20] LI, J. and CHEN, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist.* **40** 908–940. MR2985938 https://doi.org/10.1214/12-AOS993

[21] LI, Z., HAN, F. and YAO, J. (2020). Asymptotic joint distribution of extreme eigenvalues and trace of large sample covariance matrix in a generalized spiked population model. *Ann. Statist.* **48** 3138–3160. MR4185803 https://doi.org/10.1214/19-AOS1882

[22] MESTRE, X. (2008). Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Trans. Inf. Theory* **54** 5113–5129. MR2589886 https://doi.org/10.1109/TIT.2008.929938

[23] NAJIM, J. and YAO, J. (2016). Gaussian fluctuations for linear spectral statistics of large random covariance matrices. *Ann. Appl. Probab.* **26** 1837–1887. MR3513608 https://doi.org/10.1214/15-AAP1135

[24] PAN, G. (2014). Comparison between two types of large sample covariance matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 655–677. MR3189088 https://doi.org/10.1214/12-AIHP506

[25] PAN, G. M. and ZHOU, W. (2008). Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *Ann. Appl. Probab.* **18** 1232–1270. MR2418244 https://doi.org/10.1214/07-AAP477

[26] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. MR2399865

[27] PAUL, D. and AUE, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Plann. Inference* **150** 1–29. MR3206718 https://doi.org/10.1016/j.jspi.2013.09.005

[28] SCHMIDT, M., BÖHM, D., VON TÖRNE, C., STEINER, E., PUHL, A., PILCH, H., LEHR, H.-A., HENGSTLER, J. G., KÖLBL, H. et al. (2008). The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res.* **68** 5405–5413.

[29] WANG, W. and FAN, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann. Statist.* **45** 1342–1374. MR3662457 https://doi.org/10.1214/16-AOS1487

[30] YANG, Q. and PAN, G. (2017). Weighted statistic in detecting faint and sparse alternatives for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **112** 188–200. MR3646565 https://doi.org/10.1080/01621459.2015.1122602

[31] YAO, J., ZHENG, S. and BAI, Z. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis. Cambridge Series in Statistical and Probabilistic Mathematics* **39**. Cambridge Univ. Press, New York. MR3468554 https://doi.org/10.1017/CBO9781107588080

[32] YATA, K. and AOSHIMA, M. (2012). Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. *J. Multivariate Anal.* **105** 193–215. MR2877512 https://doi.org/10.1016/j.jmva.2011.09.002

[33] ZHANG, Z., ZHENG, S., PAN, G. and ZHONG, P.-S. (2022). Supplement to “Asymptotic independence of spiked eigenvalues and linear spectral statistics for large sample covariance matrices.” https://doi.org/10.1214/22-AOS2183SUPP

[34] ZHENG, S., BAI, Z. and YAO, J. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Ann. Statist.* **43** 546–591. MR3316190 https://doi.org/10.1214/14-AOS1292